

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:25:50 : Search time 6143 Seconds

(without alignments) 5503.423 Million cell updates/sec

Title: US-09-512-581b-2
Perfect score: 7193
Sequence: 1 MAHSTRNNOGKITYPPGVK.....OKGRGRPSKTPSPQPKKNV 1391

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 segs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPRO.spool/US09512581/runat_23092003_163547_5486/app_query.fasta_1.1543
-DB-BST -QFMT-fastlap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=-1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09512581 -ACGN_1_1_4667 -ernuat_23092003_163547_5486 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESOURCE -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlu:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3787	52.6	3313	11 AK045159	AK045159 Mus muscu
2	3512	48.8	2899	11 AK041682	AK041682 Mus muscu
3	3309.5	46.0	4273	11 AK086753	AK086753 Mus muscu
4	3037	42.2	4427	11 AK032384	AK032384 Mus muscu
5	2170.5	30.2	1359	11 BC032988	BC032988 Mus muscu
6	1395	18.4	926	14 CA980171	CA980171 AGENCOURT
7	1325	18.4	786	14 CA324097	CA324097 UT-M-FY0-
8	1311	18.2	779	14 CA324284	CA324284 UT-M-FY0-
9	1298.5	18.1	786	13 BU703488	BU703488 UT-M-FY0-
10	1295	18.0	782	13 BU363024	BU363024 603790485
11	1285.5	17.9	943	13 BU914579	BU914579 AGENCOURT
12	1269.5	17.6	857	14 CA987660	CA987660 AGENCOURT
13	1262.5	17.6	882	13 BU916668	BU916668 AGENCOURT
14	1239	17.2	931	13 BU914025	BU914025 AGENCOURT
15	1231	17.1	829	13 BX314916	BX314916 BX314916
16	1225	17.0	791	14 BF509252	BF509252 UT-H-B14-
17	1204.5	16.7	771	14 CB312535	CB312535 AGENCOURT
18	1189.5	16.5	748	13 BU957734	BU957734 AGENCOURT
19	1175.5	16.3	905	13 BU940978	BU940978 AGENCOURT
20	1168	16.2	692	12 BM794657	BM794657 K-EST0076
21	1162	16.2	716	14 CB522273	CB522273 UT-M-GH0-
22	1152.5	16.0	876	13 BU912768	BU912768 AGENCOURT
23	1143.5	15.9	734	10 BG256731	BG256731 602370939
24	1142	15.9	679	10 BB094813	BB094813 BB094813
25	1130	15.7	823	13 BU703952	BU703952 UT-M-FY0-
26	1128	15.7	703	14 CB521463	CB521463 UT-M-GH0-
27	1124.5	15.6	868	14 CD301710	CD301710 AGENCOURT
28	1119	15.6	658	14 CD542155	CD542155 B0238804-
29	1108.5	15.4	695	14 CA405800	CA405800 1001949 H
30	1099	15.3	659	12 BU738240	BU738240 603361679
31	1091	15.2	729	10 BG432960	BG432960 602495783
32	1083	15.1	637	13 CB132602	CB132602 K-EST0183
33	1083	15.1	643	13 BU440284	BU440284 604146435
34	1078	15.0	874	13 BU517176	BU517176 AGENCOURT
35	1076	15.0	633	14 CD542034	CD542034 B0237804-
36	1076	15.0	764	13 BU384740	BU384740 603860876
37	1071	14.9	773	9 AL705029	AL705029 DKFP686B
38	1070	14.9	804	13 BU109808	BU109808 603128343
39	1069	14.9	638	14 CD284258	CD284258 G39173.86
40	1067.5	14.8	1142	12 BM466282	BM466282 AGENCOURT
41	1066	14.8	944	10 BF142489	BF142489 601788573
42	1061.5	14.8	657	9 BE873840	BE873840 601483992
43	1061	14.8	657	9 AL846882	AL846882 AL846882
44	1061	14.8	858	13 BU736792	BU736792 AGENCOURT
45	1056.5	14.7	624	14 CB215704	CB215704 NISC_np09

ALIGNMENTS

RESULT 1
LOCUS AK045159 3313 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
enriched library, clone: B130042B12 Product: 49310.1.2
(ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
sequence.
ACCESSION AK045159
VERSION AK045159.1 GI:26337114
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schmitt, L. M., Staahl, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Morichioni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Wittaker, C., Wilmink, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 3313)
PUBMED	12108561
REFERENCE	Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shikata, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hnako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/. Location/Qualifiers
FEATURES	1. .3313
SOURCE	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM,DB:BI30042B12"
	/db_xref="taxon:10090"
	/clone="BI30042B12"
	/tissue_type="parthenogenote"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="9.5 days embryo"
CDS	<1..2476
	/note="unnamed protein product; 49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (PRAGENT) homolog (Homo sapiens) (SPR1Q96KY4, evidence: FASTV, 92.3%ID, 100%length, match=1430) putative"
	/protein_start="2"
	/protein_id="BAC32242.1"
	/db_xref="GI:26337115"
	/translation="VNKSIDGTADDEDEGEVTPDQATPAGIELIKVLSTPFIHSFAE TPESLACLMDEDEKVAEALQIFKNTGSEIEDPFRIRALLPVLHKKKKPPROA KVAIHCHAIFFSKETFOAFIEPLHNSLSDPSNDEHITPLVTGHTALAPQFAP KLSLVAIFVIFDLMDNDRLPGKRTYKLVDPDEVSPETWAKIQAIKMWVLLGMKN HSQSGSTLTLTLTIHSDGLDEOGKISRPMSRLTAAGSAIVKLAQPCYHEIIT LEOYQCALAINBECYQVROVFPKLGKISGRRLPLEYVATICALAKDPYERRAHA RQCLVKITVYRELVKQHAVALSKSLILEYVVPYTHLADPQVYKQDIEQGLD VKRCIMVPELLEIMAKNNNSHAFIRKRVENIKOTKQAQGDPTKMKKNTYTCDVAM IIMSKSTTYSLESPKDPVLPAREFTOPDNFNTKNTLPPKMSFTPGKPTKATVIG AVNKPUSASAKQSQVTSSRMETVSNASSNPSPGKIKGRLDSSSEHDENDYVMS SPLPGKRSKREDPDLVRSLEKPRSKKAPVNDPEKILGMDLTKLVOEOKFGSOR GRRKGRTAQSDSQEOMPEERKHEKLELLENDQNSPKKGRGRRPPLGGSKSEEP TMTSKKGNKKLVPPVDDDEEERDIGNTEHRSKSKOHRTSKRAQOVRRESQTSV ESNQSPQKGRGRSPKAPSPQPKTRVGRSKQVATKENDSSSEMDVLDASSPVDD TPOBGADEEDISGVNRRSSKERR"
	polyA_signal
	polyA_site
BASE COUNT	1066 a 709 c 732 g 806 t
ORIGIN	3313
	/note="putative"
	3313
Alignment Scores:	7.51e-275
Pred. No.:	3787.00
Score:	97.00%
Percent Similarity:	95.31%
Best Local Similarity:	52.65%
Query Match:	11
DB:	Gaps: 0
US-09-512-581b-2 (1-1391) x AK045159 (1-3313)	
QY	625 ValAsnlySerIleaspGlyThrAlaaspAspGluuAspGluGlyValProThrAspGln 644

|||||
Db 2 GTGAATAATCATGATGGAACACAGATGATGAGAGGTTCACACTATCA 61
Oy 645 AAlaIleargAlaGlyLeuGluLeuLeuLysValLeuSerPheThiHisProIleSerPhe 664
Db 62 GCTATCAGAGGGAGGCTTGAGCTGCTGAAGGTGTGTCTTTACCCACTATTCCTTT 121
Oy 665 HisSerIaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysVal 684
Db 122 CACTGTGCTGAGAGCTTTGAGTCCCTTCTGCTTGTCTGAGATGATGATGATGAGAGG 181
Oy 685 AAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPhePro 704
Db 182 GCAGAGAGCTGCACATCAATTTTCAAAAAACAGAGAACAAATGGAAGAGATTTCCCT 241
Oy 705 HisIleArgSerAlaLeuLeuProValLeuHisSHLysSerLysGlyProProArg 724
Db 242 CACATCAGATCAGCCCTTCTCTGCTACTACATCAAAATCCAAAAAGGACCCCTCG 301
Oy 725 GlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPhe 744
Db 302 CAGGCCAAATACGCCATTCATTTTTCATGCCATATTTTCTAGTAAGAAACCCAGTTT 361
Oy 745 AlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIle 764
Db 362 GCACAGATATTTGAGCCCTGCATAAAGCCTAGATCCAGCACCTGGAGCATCTTATA 421
Oy 765 ThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaPro 784
Db 422 ACCCTCTGTGCTACTATTGTGCATATTTCTCTTCGACCTGATCAGTCCCTCTCT 481
Oy 785 TrpLysSerTrpValAlaThrPheIleValLysAspLeuMetMetAsnAspArgLeuPro 804
Db 482 CTGAAGTCTTTGGTGGCACTTTCATTTGTGAAGGCCCTCTCATATATGACCGGCTTCCA 541
Oy 805 GlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMetVal 824
Db 542 GGAATAAAGACAACTAACTTTGGGTCCAGATGAGAGAGTCTCACTGAGACAAATGGTC 601
Oy 825 LysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnHisSer 844
Db 602 AAAATTCAGGCTATTAAATGATGTGTCATGCTGATCTGCAATGCAATGATATCATAGT 661
Oy 845 LysSerGlyThrSerThrLeuArgLeuLeuThrTrpIleLeuHisSerAspGlyAspLeu 864
Db 662 AAGTCAGGAACCTCCACCCCTCAGATTACTAACACGATACGATGATGATGGGATTTG 721
Oy 865 ThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaIaGlySer 884
Db 722 ACAGAACAAAGGAAATTAGTAACCAAGATATGTCAGCCCTGAGACTTGTGCTGGAGT 781
Oy 885 AlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
Db 782 GCTATTGTGAAGCTGGGCACAGAGCCCTGTTACCCAGAGATCATTCACCTGGAGCAGTAC 841
Oy 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
Db 842 CAGCTGTGTCATTAAGCATATGATAGTGTATTATCAAGTCAGGCAAGTGTTCCTCAG 901
Oy 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluLysMetAlaIleCysAla 944
Db 902 AAACCTTCACAAAGGCTTTCCGCTTACGGCTTCCCTTGAAGTACATGGCCATCTGTCT 961
Oy 945 LeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLys 964
Db 962 CTTTGTGCAAAAGACCTGTGTAAGAGAGGCGAGCCCTGCTAGCACTGTCTGTGAAG 1021
Oy 965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984
Db 1022 AACATCACTGTGAGAGGAGTACTCTGAAGCAGCTGAGCTGTTAGTGAATAATTTATG 1081
Oy 985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
|||||

Db 1082 TCTCTTCAACGAGTATGTTGCTCCATATACATTCACCTTTTGGCAGATGACCCAGAT 1141
Oy 1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheVal 1024
Db 1142 TATGTCAAGATACAGATATATGACAACTTAAAGATGTGAAGAATGCCCTTTGGTTGTT 1201
Oy 1025 LeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044
Db 1202 CTGGAGATATATGATGGCTAAATAATGAACAAACACACCATGATTTATCAGAAAAATGGTA 1261
Oy 1045 GlnAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
Db 1262 GAAATATATTAACACACAAATAAGTCTCAGAGCACAGATGATACAAATAATGAAAAA 1321
Oy 1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
Db 1322 TTGTACCCGTGTGTGATGTTCATGAACATCATCATGTCCAAAGCACCCAGTACAGC 1381
Oy 1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
Db 1382 CTGGAGTCTCCTAAGAGACCCCGTGTGCTGACGCTGGTTTTCACCCAGCCTGACAAAGAT 1441
Oy 1105 PheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124
Db 1442 TTTAGTAAACACCAAAATATACCTGCTCCAGAAATGAATCATTTTTCACCTCTGAAAA 1501
Oy 1125 ProLysThrThrAsnValLeuGluValAlaLysLysProLeuSerSerAlaGlyLysGln 1144
Db 1502 CCTAAACAGCCAAATGTTCTCGACACAGTAAATTAAGCCATTTCATCAGCAGCAACAG 1561
Oy 1145 SerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSerSerAsnPro 1164
Db 1562 TCTCAGACCAATATCATCAAGATGGAACCTGAGACCAACGCAAGCAGCAGCTCAACCA 1621
Oy 1165 SerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsn 1184
Db 1622 AGCTCTCTCGAAGGATCAAGGGGAGGCTTGTATCTCTGAAATGATGCACAGTGAAT 1681
Oy 1185 GluAspTyrThrMetSerSerProLeuProGluLysLysSerAspLysArgAspAspSer 1204
Db 1682 GAAAGATATACATGTCTTACCTTGGCAGGAAAAAAGACAAAGAAAGAACCT 1741
Oy 1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrProValThrGlu 1224
Db 1742 GATCTTGTAGGTCTGAGTGTGAGAGAACCTAGAACTCGGAAAAAACACCTCTCAGAC 1801
Oy 1225 GlnGluGluLysLeuGlyMetAspAspLeuThrThrLysLeuValGlnGluGlnLysProLys 1244
Db 1802 CCTGAAGAAATTAGTATGATGATGACCTAACTAAGTTGGTACAGAACAGAAACCTAAA 1861
Oy 1245 GlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrp 1264
Db 1862 GCCAGTCAAGCGCGCGGAAAAAGAGCCGTACAGCTTACAGCTCAGCAGCAGAGTGG 1921
Oy 1265 ProGluGluLysArgLeuLysGluLysAspIleLeuGlnGluAsnGluLysGluGlnAsnSerPro 1284
Db 1922 CCTGAGAGAAAGGACACAAAGAGAGAGCTCTCGGAAAAATGAGAGTACAGACAGACCCA 1981
Oy 1285 ProLysLysGlyLysArgGlyValArgProProLysProLeuGlyGlyGlyThrProLysGlu 1304
Db 1982 CCAAAAAAGGCGCAAAAGAGGAGCGCACCAAAACCTTGGTGGGGGAGACATCGAAGAA 2041
Oy 1305 GluProThrMetLysThrSerLysGlySerLysLysSerGlyProProAlaPro 1324
Db 2042 GAGCCCAATGAAATAATCCAAAGAAAGAAACAAAGAAAACTTGTACTCTCTGTAGTA 2101
Oy 1325 GlnGluGlnGluGlnGluGluArgGlnSerGlnLysAsnThrGlnGlnLysSerLysSerLys 1344
Db 2102 GACGACGATGAAGAAAGAAAGAAACAATTTGGAACACAGAAACATTAAGTCAAAAAAGCAA 2161
Oy 1345 GlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIle 1364
Db 2162 CAGCACCCGACATCAAAAGAGAGCACAAAGAGAGAGCAATCTCTGAAAAACAAGTCACTT 2221
|||||

QY 1365 GluSerThrInsSerThrProGlnIlySGlyArgGlyArgProSerIyThrProSerPro 1384
 |||||||
 Db 2222 GAATCCACAGACGATCCACACACAGAAAGACGAGAAACACATCAAAAGACCATCCCA 2281
 QY 1385 SerGlnProIySlySasnVal 1391
 |||||||
 Db 2282 TCACAACCCCCCAAAAAATC 2302
 RESULT 2
 AK041682
 LOCUS DEFINITION 2899 bp mRNA linear HTC 05-DEC-2002
 Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
 library, clone:A630029M15 product:49J10.1.2 (ANDROGEN-INDUCED
 PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
 (FRAMMENT) homolog [Homo sapiens], full insert sequence.
 AK041682
 VERSION AK041682.1 GI:26334660
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL Prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE 20499374
 PUBMED 11042155
 REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishize, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiyagi, K.,
 Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Tameda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, D.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE
 AUTHORS Arakawa, T., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
 Saito, T., Okazaki, Y., Gojodori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L. M., Straub, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
 Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,
 and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660

PUBMED 11217851
 REFERENCE
 AUTHORS 5
 TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research
 JOURNAL Group Phase I & II Team.
 MEDLINE Analysis of the mouse transcriptome based on functional annotation
 PUBMED of 60,770 full-length cDNAs
 REFERENCE Nature 420, 563-573 (2002)
 AUTHORS 6 (bases 1 to 2899)
 Pukuda, S., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 COMMENT
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
 URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of genome exploration research group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: http://genome-gsc.riken.go.jp/
 URL: http://fantom-gsc.riken.go.jp/
 location/Qualifiers
 1. .2899
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM,DB:A630029M15"
 /db_xref="taxon:10090"
 /clone="A630029M15"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 117. .2222
 /note="unnamed protein product; 49J10.1.2
 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF
 ASSOCIATED PROTEIN, ISOFORM 2) (FRAMMENT) homolog [Homo
 sapiens] (SPR1096KV4, evidence: FASIV, 92.3%ID,
 100%length, match=1430)
 putative"
 /codon_start=1
 /protein_id="BAC31031.1"
 /db_xref="GI:26334661"
 /translation="MTITRPLPPGKAQPMKFTQVLEDEKIRKQLEALVSPSC
 KQAGCYREITKILGNPOTNPFLMKIRLIRIVPHVIDESTIALKROVKSISG
 TADDEGCVPTDAIRAGELIVLSFTHPSRSPSTPSILACIKMDKRYAABAL
 QIRKNGSKIEDPFPRIKALVYLHKKSKGPPROAKTIAHCTHAFVSKETQPAOI
 FEPLHKSLSDBSNELHITPLVTIGHIALAPDPAFLKSLVTFIVKDLMDRLDG
 KKTITKLWPDPEVSPEPWVIAIKRWMLKNNHSGSTSLRLTILHSODG
 LTFQGIKSPDKMSRLRLAAGSAIVKLAQECYIEITLQYICALAINDECYQAV
 FAQIKHKGSRRLPLEYMAIKCAKDKPKERBARAROCIVNITVREYKOHAAV
 SEKILSLIPYVYVPIHIALHPDVYKODIDQIDQVCEGLPVEITIMARKENSH
 APTRKAVENIKOTKDAQGPDDTKMNEKLTVCVYANNTIKSKSTTSLSPKQPVLP
 RFTTQDPKNSNTKNYLPPEMKSPFTPGKPTANVIGAVKPLSSAGKOSQTSRME
 TVSNASSNPSSNGRIKGRLDSEMDHSENEDEYTPSSPLPKKSDKREDPDLVRVC

BASE COUNT 868 a 582 c 617 g 832 t
ORIGIN LVGRMRLIVLVIPAL*
Alignment Scores:
pred. No.: 3,51e-254 Length: 2899
Score: 3512.00 Matches: 683
Percent Similarity: 98.99% Conservative: 2
Best Local Similarity: 98.70% Mismatches: 7
Query Match: 48.83% Indels: 0
DB: 11 Gaps: 0
US-09-512-581b-2 (1-1391) x AK041682 (1-2899)
QY 517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
DB 78 CAGACAGATGCCAGTGTCTCAAGCCCAATTTTCTAAAGTATGTTATCCACAAATTTTG 137
QY 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAsp 556
DB 138 CCAGATCCCTGCTAAGGCCCAAGACCTTCATGAAGAAGTTCCACACAAGTGTGGAAAGATGAT 197
QY 557 GluLysIleArgLysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnAla 576
DB 198 GATGAATAATGGAACAGCTGGAAGCCCTTGTTAGTCCCAATGTTCTGCAAGCAGGCT 257
QY 577 GluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnPro 596
DB 258 GAAGGCTGTGCGTGAATTAAGAGTTGGGCAACCTTAAGCAGGCTTCAAAACCT 317
QY 597 PheLeuGluMetCileLysPheLeuLeuGlnArgIleAlaProValHisIleAspThrGlu 616
DB 318 TCTCTGGAATATCATCAAGTTCTCTTGAGAGATAGCTCTGTACATAGACACTGAA 377
QY 617 SerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGlu 636
DB 378 TCCATAGTGTCTTATCAAGCAAGTGAATTAATCAATAGATGGAACACAGATGATGAG 437
QY 637 AspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGlnLeuLeuLysValLeu 656
DB 438 GATGAGGGTGTCCAACTGATCAGGTATCAGGGCAGGCTTGAGTGTGAAGTGTGCTG 497
QY 657 SerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
DB 498 TCTTTTACGACCCCTATCTCTTTCACCTGCTGAGACGTTGAGTCCCTTGCGCTGT 557
QY 677 LeuLysMetAspAspGluLysValAlaGlnAlaIleLeuGlnIlePheLysAsnThrGly 696
DB 558 CTGAAGATGATGATGAGAAGGTGGCAGAAGCTGCACATAAATTTTCAAAAACACAGGA 617
QY 697 SerLysIleGluGlnAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHis 716
DB 618 AGCAAAATTTGAAGAGATTTCCCTCACAATCAGATCAGCCTTGCTTCTACTACTACATC 677
QY 717 LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
DB 678 AAATCCAAAAAGGACCACTCGGACGAGGCAATACGCCATCATGTGATTCATGACATA 737
QY 737 PheSerSerLysGlnThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
DB 738 TTTTCTAGTAAGAAACCCAGTTTGACAGATATTGAGCCTCTGCAATAAAGCCCTAGAT 797
QY 757 ProSerAsnLeuGlnHisIleLeuIleThrProLeuValIleGlyHisIleAlaLeuLeu 776
DB 798 CCAACCAACTGTGAGACATCTTATTAACCCCTGTCGACATATGGTCATATTCCTCTTC 857
QY 777 AlaProAspGlnPheAlaAlaProThrLysSerTyrValAlaIleThrPheIleValLysAsp 796
DB 858 GCACCTGATCAAGTGTGCTGCTCTGGAAGCTTTGGTGGAACCTTTCATGTGGAAGGAC 917
QY 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThrValProAspGlu 816
DB 918 CTCCTCATGAAATGACCGGCTTCCAGGAATAAAGACAACTAAGCTTTGGGTTCCAGATGAG 977

QY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
DB 978 GAAGTCTCACCTGAGACATGCTCAAAATTCAGGCTATTAATATGATGTTGATGAGGCTA 1037
QY 837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThr 856
DB 1038 CTTGGAAATGAATAATATCAGATGAAGTCCAGAACTTCCACCCTCAGATTACTTAACAAAG 1097
QY 857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
DB 1098 ATACTGCATGATGATGGGATTTTGACAGAACAGGGAATTAAGTAACCAATATGATGA 1157
QY 877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnIleProCysTyrHis 896
DB 1158 CGCCTGAGACCTTGCTGCTGGAGTGTATTTGAAGCTGGACAGGAGCCCTGTTAACCAAC 1217
QY 897 GluIleIleThrLeuGluGlnIleThrGlnLeuCysAlaLeuAlaIleAsnAspLysCysTyr 916
DB 1218 GAGATCATTTACACTGGAGCAGTACCAAGCTGTGTCATTAAGCATCAATGATGAGTGTAT 1277
QY 917 GluValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgPro 936
DB 1278 CAGTTCAGGACAGGTGTCGCTCAGAACTTCACAAAGGCTTCCCGCTTACGCTTCCC 1337
QY 937 LeuGluIleThrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
DB 1338 CTTGAGTACATGGCCATCTGTGCTCTTGCCAAAGACCTGTGAAGAGAGGAGGCC 1397
QY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluIleLysGlnHis 976
DB 1398 CATGCTAGACACTGTGGTGAAGAACACTGTGAGGAGGAGTACTGTAAGCAGCAT 1457
QY 977 AlaAlaValSerGluLysLeuSerLeuLeuProGluThrValValProThrThrIle 996
DB 1458 GCAGCTGTAGGAAATATATGTCTCTTCAACAGATATGTTGTCATATCAATAT 1517
QY 997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
DB 1518 CACCTTTTGGCAGATAGCCACAGATATATGCAAAAGTACAGAGATATTGAACAACTTAAGAT 1577
QY 1017 ValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsnGlnAsnAsnSer 1036
DB 1578 GTGAAGAATAGCTTTGGTTGTTCTTGAGATATTGATGGCTTAAATAAGTAAACAACAGC 1637
QY 1037 HisAlaPheIleArgLysMetValGlnAsnIleLysGlnThrLysAspAlaGlnIlePro 1056
DB 1638 CATGCATTTATCAGAAATAATGATAGAAATATTAAACAGCAAAAGATGCTCAAGGACCA 1697
QY 1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
DB 1698 GATGATACAAAATAATGATTAATAATTTGTACACGCTGTGATGTTGCCATGAACATCATC 1757
QY 1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
DB 1758 ATGTCCAAAGACACACAGTACAGCCTTGAGTGTCTCTTAAGGACCCGTGCTGCACACTGG 1817
QY 1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMet 1116
DB 1818 TTTTTCACCCACCTCGACAGAAATTTTGTACACCAAAATTTACTGCTTCCAAATATG 1877
QY 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
DB 1878 AAATCATTTTTCACCTCGGAAACCTTAACACAGCAATGTTCTCGAGAGATTAAATAG 1937
QY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerArgMetGluThrValSer 1156
DB 1938 CCACCTTTCATGACGAGGAAACAGTCTCAGACCAATATCATCAAGATGGAACCTGTGAGC 1997
QY 1157 AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSer 1176
DB 1998 AACGCAAGCAGCAGCTCCAAACCAAGCTCTCTGGAAGATTCAGGGAGGCTTGTATAGC 2057

Oy 1177 SerGluMetAspHisSerGluAsnGluAspTyrThrMetSerProLeuProGlyLys 1196
 Db 2058 TCTGAATGATGCACAGTAAGAAATGAAATGATATACATGCTTCACCTTCACGAGNAAA 2117
 Oy 1197 LysSerAspLysArgAspAspSerAspLeuValArg 1208
 Db 2118 AAAAGTGCAAGAGAGAACCCCTGATCTTGTAAAG 2153

RESULT 3
 AK086753
 LOCUS
 DEFINITION Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:0930049H23 product:49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert sequence.

ACCESSION AK086753
 VERSION AK086753.1 GI:26103739
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, S., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiyake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Kureda, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

4
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojodori, T., Bono, H., Kasukawa, T., Saito, R., Kaotaka, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Knehl, P., Lewis, S., Matsuo, Y., Mikado, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stanbail, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Kono, H., Baldarelli, R., Barish, G., Blake, J., Boffelli, D., Boljuga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Butt, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamlya, M., Lee, N. H., Lyons, P., Marchionni, J., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Boriss, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 JOURNAL
 MEDLINE
 PUBMED 21085660
 11217851

REFERENCE 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 4273)

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Yoshida, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/.
 URL: http://fantom.gsc.riken.go.jp/.
 Location/Qualifiers

FEATURES
 source
 1..4273
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM,DB:D930049H23"
 /db_xref="taxon:10090"
 /clone="D930049H23"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="15 days embryo"
 <1..2245
 /note="49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog (Homo sapiens) (SPTR1096KV4, evidence: FASTV, 92.3%ID, 100%length, match=1430) putative"

misc_feature
 <1..2245
 /note="49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog (Homo sapiens) (SPTR1096KV4, evidence: FASTV, 92.3%ID, 100%length, match=1430) putative"

BASE COUNT
 1363 a 840 c 899 g 1171 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.06e-238 Length: 4273
 Score: 3309.50 Matches: 652
 Percent Similarity: 96.24% Conservative: 14
 Best Local Similarity: 94.22% Mismatches: 23
 Query Match: 46.01% Indels: 5
 Gaps: 1

US-09-512-581b-2 (1-1391) x AK086753 (1-4273)

Oy 702 AspheProHisIleArgSerAlaLeuLeuProValLeuHisIstLysSer-LysLysG 721
 Db 1 GATTTCCCTGCACATCAGATGAGCCCTTCCTCTACTACATCAACAATCAAAAAAG 60

Oy 721 yProPArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLysG 741
 Db 61 ACCACCTGGGAGGCGCAATGAGCCATGCATGATGATGATTTCTAGTAAGA 120

Oy 741 uThrgInpheAlaInJlePheglUpProleuHIsLySserLeuAspProSerAsnLeuGl 761
 |||||
 Db 121 AACCCAGTTGGACAGATATTGGAGCCTGTCATATAAACCTTAGATCCAGCAAGCACTGGA 180
 Oy 761 uHIsLeuIleThrProLeuValThrIleGIyHIsIleAlaLeuLeuAlaProAspGlnPh 781
 |||||
 Db 181 GCATCTTAATACCCCTCTGTGCATCTATGTGCATATTGTCTCTTCGCACTGATCATGATT 240
 Oy 781 eAlaAlaProTrpLySserTrpValAlaThrPheIleValLySAspLeuLeuMetAsnAs 801
 |||||
 Db 241 CCGTCTCCCTGAGAGTCTTTGGTGCACCTTTCATGTGGAAGGACCTCTCATGTGATGA 300
 Oy 801 pArgLeuProGIlyLySThrThrLySLeuTrpValProAspGIuGIuValSerProGI 821
 |||||
 Db 301 CCGGCTCCAGGAAAAAAGACAACATAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCTTA 360
 Oy 821 uThrMetValLySIIeGIAlaIleLySMeMetValArgTrpLeuLeuGIyMetLySAs 841
 |||||
 Db 361 GACAAATGGTCAAAATTCAGGCTAATTAAATGATGGTTCGATGGCTACTTGGAAATGAAGA 420
 Oy 841 nAsnHIsSerLySserGIyThrSerThrLeuArgLeuLeuThrThrIleLeuHIsSerAs 861
 |||||
 Db 421 TTAATCACTAGTAAGTCAGAACTCCACCTCAGATTACTAACAACGATACCTGCATATGTA 480
 Oy 861 pGIyAspLeuThrGIuGIuGIyLySIIeSerLySProAspMetSerArgLeuArgLeuAl 881
 |||||
 Db 481 TGGGATTTGACAGAAACAAGGAAAAATAGTAACAGATATGTCAACGGCTGAGACTTGC 540
 Oy 881 aAlaGIySerAlaIleValLySLeuAlaGIuGIuProCysTyRHisGIuIleIleThrIe 901
 |||||
 Db 541 TCCTGGAGAGTCATTATGGTAACCTGCACAGAGGCCCTGTACCAAGATATATTCACACT 600
 Oy 901 uGIuGIuThrGIuLeuCysAlaIleuAlaIleAsnAspGIuCysTyRGIuValArgGlnTa 921
 |||||
 Db 601 GGAGCAGATCCAGCTGTGTGCATTTAGCCATCAATGATGATGATTTATCAAGTCAAGGCA 660
 Oy 921 lPheAlaGIuLySLeuHIsLySGLyLeuSerArgLeuArgLeu - ProLeuGIuTyMetA 941
 |||||
 Db 661 GTTCCTCGAAGAACTTCAACAAAGGCCCTTCCCGCTTAGAGGCTTCCCTTGGAGTACATG 720
 Oy 941 lAlleCysAlaLeuCysAlaLySAspProValLySGluArgArgAlaHIsAlaArgGlnC 961
 |||||
 Db 721 CCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAGAGAGGGAGGCCATCTGTACACACT 780
 Oy 961 ySLeuValLySAsnIleAsnValArgArgGIuTyRLeuLySGlnHIsAlaIleValSerG 981
 |||||
 Db 781 GCTTGTGAGAGAACATCCC -GTGAGAGAGGAGTACCTGTAAGCAGCTGCAGCTGTATAGT 839
 Oy 981 lUlySLeuLeuSerLeuLeuProGIuTyRValValProTyRThrIleHIsLeuLeuAlaH 1001
 |||||
 Db 840 AAAAATTAATTGTCTCTTACCAAGATATTGTTCATATACAAATTCACCTTTGGGCAC 899
 Oy 1001 lAspProAspTrpValLySValGIuAspIIeGIuGIuLeuLySAspValLySGluCysL 1021
 |||||
 Db 900 ATGACCCAGATTATGTCAAAAGTACAGGATATTGAACAACCTTAAAGTGTGAAGAAGATGCC 959
 Oy 1021 eutRrPheValLeuGIuIleLeuMetAlaLySAsnGIuAsnAsnSerHIsAlaPheIleA 1041
 |||||
 Db 960 TTTGGTTTGTCTGAGAGATATTGATGGCTAAATAAAGAAACAACAGCCCTGCATTTATCA 1019
 Oy 1041 rGlySMeMetValGIuAsnIleLySGlnThrLySAspAlaGIuGIuProAspAspAlaLyS 1061
 |||||
 Db 1020 GAAAAATGGTAGAAAAATATTAAACAGACAAAGATGCTCAAGAGACAGATGATACAAAAA 1079
 Oy 1061 eTAsnGIuLySLeuTyRThrValLySAspValAlaMetAsnIleIleMetSerLySserT 1081
 |||||
 Db 1080 TGAATGAATAATGTACACCGGTGTGATGTTCATGAACATCAATCAATGATCCAAAGAGA 1139
 Oy 1081 hTrThrTySerLeuGIuSerProLySAspProValLeuProAlaArgPhePheThrGlnP 1101
 |||||
 Db 1140 CCACGTACAGCTCGAGTCTCTTAAGAGACCCCGTGTGCTGCAGCTGGTTTTCACCCACAC 1199
 Oy 1101 roAspLySAsnPheserAsnThrLySAsnTyRLeuProProGIuMetLySserPhePheT 1121

Db 1200 CTGACAGAAATTTTAACTACACCAAAATTAACCTGCTCCAGAAATGAATATATTTC 1259
 Oy 1121 hTrProGIyLySProLySThrThrAsnValLeuGIuValAlaLySAsnLySProLeuSerSera 1141
 |||||
 Db 1260 CTCCTGGAAACCTTAAACAGCCAAATGTTCTCGAGCAGTATATAAGCCACTTTCACAG 1319
 Oy 1141 lAGlyLySGlnSerGIuThrLySserSerArgMetGIuThrValSerAlaSerSers 1161
 |||||
 Db 1320 CAGGCAAAAGTCTCAGACCAATATCAAGAAATGGAAACTGTAGCAACGCAAGCAGCA 1379
 Oy 1161 eSerAsnProSerSerProGIyArgIleLySGIyArgLeuAspSerSerGIuMetAspH 1181
 |||||
 Db 1380 GCTCCACCCCAAGCTCTCTCGAAGATAG -GGAGGCTGTATACCTTGAATATGAGATC 1438
 Oy 1181 lASerGIuAsnGIuAspTyRThrMetSerSerProLeuProGIyLySAspLySA 1201
 |||||
 Db 1439 ACAGTGAATAATGATATATACAAATGCTTTCACCTTTCAGCAAAAAAAAGTGCACAGA 1498
 Oy 1201 rGAspAspSerAspLeuValArgSerGIuLeuGIuLySProArgGIyArgLySThrP 1221
 |||||
 Db 1499 GAAAGACCTTATCTTTGTAAGTCTGACTTGGAAAGCTTGAAGTGGAAAAAAGCAC 1558
 Oy 1221 roValThrGIuGIuGIuGIuLySLeuGIyMetAspAspLeuThrLySLeuValGIuGIu 1241
 |||||
 Db 1559 CTGTCAAGACCCCTGAAGAGAAATTAGTATGATGACTTAAGTTGGTACAGAAC 1618
 Oy 1241 lUlySProLySGIySerGIuArgSerArgLySArgIyHIsThrAlaSerGIuSerAspG 1261
 |||||
 Db 1619 AGAAACCTTAAAGGAGTGCAGGAGCCGGAAGAGGCGCTTCACCTTCAGCTCAGACG 1678
 Oy 1261 lUGIuGIuThrProGIuGIuLySArgLeuLySGluAspIIeLeuGIuAsnGIuAspGIu 1281
 |||||
 Db 1679 ACACAGAGTGGCTTGAGGAGAAAGAGCACAAAGAGAGTCTTGAAAAATGAGATGAC 1738
 Oy 1281 lAsnSerProProLySLySGIySArgGIyArgProProLySProLeuGIyGIyGIyT 1301
 |||||
 Db 1739 ACAGACAGCCCAAAAGAGGCAAAAGAGGAGGAGCCACCAACCTTGTGTGGGGGA 1798
 Oy 1301 hTrProLySGluGIuProThrMetLySThrSerLySLySGIySerLySserGIyP 1321
 |||||
 Db 1799 CATCGAAGGAAGAGCCCAAAATGAACATCCAAAGAAAGAAACAAAGAAAAAATCTGTAC 1858
 Oy 1321 roProAlaProGIuGIuGIuGIuGIuGIuGIuArgGIuSerGIyAsnThrGIuGIuLyS 1341
 |||||
 Db 1859 CTCCTGTACTACACGACGATGAGAACAAAGAAAGCAAAATTTGAAACACAGAACATAACT 1918
 Oy 1341 eRlySserLySGlnHIsArgValSerArgArgAlaGIuGIuArgAlaGIuSerProGIu 1361
 |||||
 Db 1919 CAAAAAGCAAAACAGACACGACATCAAAAGAGAGCACAAACAGACAGAAATCTCTGAAA 1978
 Oy 1361 eSerAlaIleGIuSerThrGIuSerThrProGIuLySGIyArgGIyArgProSerLyST 1381
 |||||
 Db 1979 CAAGTCAAGTTAAATCCACACAGTCCACACCAACAAAGAGAGAGGAAGACATCAAAAG 2038
 Oy 1381 hTrProSerProSerGIu---ProLySAsn 1390
 |||||
 Db 2039 CACCATCACATCACACACCCCCCAAAAAAAT 2070
 RESULT 4
 AK032384 4427 bp mRNA linear HTC 05-DEC-2002
 LOCUS AK032384
 DEFINITION Mus musculus adult male olfactory brain cDNA, Riken full-length
 enriched library, clone:6430510p15 product:49310.1.2
 (ANDROBEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
 PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
 sequence.
 ACCESSION AK032384
 VERSION AK032384.1 GI:26082814
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible]

COMMENT	FEATURES
Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Sueniwa-cho, Tsurumi-Ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
Please visit our web site for further details.	
URL: http://genome.gsc.riken.go.jp/	
URL: http://phantom.gsc.riken.go.jp/.	
Location/Qualifiers	
1. .4427	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C57BL/6J"	
/db_xref="FANTOM DB:6430530P15"	
/db_xref="taxon:10090"	
/clone="6430530P15"	
/sex="male"	
/tissue_type="olfactory brain"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="adult"	
<1. .2062	
/note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHOTOF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog (Homo sapiens) (SPTRIQ96KV4, evidence: FASTV, 92.3%ID, 100%length, match=1430)	
polyA_signal	
4406. .4411	
putative"	
polyA_site	
4427	
/note="putative"	
BASE COUNT	
1423 a 832 c 919 g 1253 t	
ORIGIN	
Alignment Scores:	
Score: No.: 3.96e-218 Length: 4427	
Percent Similarity: 3037.00 Matches: 591	
Best Local Similarity: 95.888 Conservative: 14	
Query Match: 42.22% Mismatches: 24	
DB: 11 Gaps: 1	
US-09-512-581B-2 (1-1391) x AK032384 (1-4427)	
OY 761 G1UH1SLEU1EThrProLeuValThrIleGlyHisIleAlaLeuAlaProaspGln 780	
Db 3 GACCACTCTATAAACCCTCTGCTCACTATGCTCATATGCTCTTCGACCTGATCAG 62	
OY 781 PheAlaIleProThrPyrSerTrpValaIleThrPheIleValIleValAspLeuMetAsn 800	
Db 63 TTGGCTGCTCTCTGAGCTTTGGTGGCACTTTCATTGTAAGGACCTCTCATGAAAT 122	
OY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProaspGluGluValSerPro 820	
Db 123 GACCGGCTTCAGAAAGAAAGACACACTAAGCTTGGTTCAGATGAGGAAGCTCCACCT 182	
OY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840	
Db 183 GAGACATAGTGTCAAAATTTGAGCTTATTAATATGATGAGTTCGATGGCTAATTGCAATGAC 242	
OY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860	
Db 243 AATAATCAGAGTAAAGTCAGAACTTCACACCTCAGATTACTAACAAGATCTGCAATG 302	
OY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProaspMetSerArgLeuArgLeu 880	
Db 303 GATGGGATTTGACAGAACAGGAAATTTGTAACAACAGATATGTGACGCTGAGACCTT 362	

QY 881 AAlaAGlySerAlaIleValLysLeuAlaGlnGluProCysThrHisGlnIleIleThr 900
 |||||
 Db 363 GCTGCTGGAGAGTGTATTGTGAAGCTGGACACAGACCCTGTTACCACAGATCATTAACA 422
 QY 901 LeuGluGlnIleThrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
 |||||
 Db 423 CTGGAGCAGTACACCTGTGTGCATTAGCCATCAATGATGAGTGTATCAAGTCAGACGAG 482
 QY 921 ValPheAlaGlnIleLeuHisIleLysGlnLeuSerArgLeuArgLeuProLeuGluTyrMet 940
 |||||
 Db 483 GGTTCGTGCACAAACTTCACAAAGGCCCTTCCCTTACGGCTTCCCTTACGATGACATG 542
 QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGlnArgArgAlaHisAlaArgGln 960
 |||||
 Db 543 GCCATCTGTGCTCTTGTGTCACAAAGCCTGTGAAAGAGAGGAGGCCATCTGTACAGC 602
 QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaValSer 980
 |||||
 Db 603 TGCCTGTGTCAAAACATCACTGTGAGGAGGAGTACCTGACAGCAGCATCCTGTACT 662
 QY 981 GlyLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuAla 1000
 |||||
 Db 663 GAAAAATTAATCTCTCTTACCAGATGATGTTGTCATATACAAATTCACCTTTGGCA 722
 QY 1001 HisAspProAspTyrValLysValGlnAspIleGlnGlnLeuLysAspValLysGluCys 1020
 |||||
 Db 723 CATGACCCAGATTATGTCAAGTACAGATATTGAACACTTAAGATGTGAAGATGC 782
 QY 1021 LeuTyrPheValLeuGluIleLeuMetAlaLysAsnGlnAsnSerHisAlaPheIle 1040
 |||||
 Db 783 CTTTGGTTGTTCTGTGAGATATTGATGGCTAAATAAGAACACGCCATCATTTATTC 842
 QY 1041 ArgLysMetValGlnAsnIleLysGlnIleThrLysAspAlaGlnGlyProAspAlaLys 1060
 |||||
 Db 843 ACAAAAGTGTAGAAATTAATTAACAGACAAAGAGTGCACAGACCAATATACAA 902
 QY 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
 |||||
 Db 903 ATGAATGAAAAATTTGTACACCGTGTGTGATGTTGCCATGACATCATCTGTCCAAAGAC 962
 QY 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
 |||||
 Db 963 ACCACGTACAGCCTGTGAGTCTCTCAAGACCCCGTGTCTCCAGCTCGGTTTTCACCCAG 1022
 QY 1101 ProAspLysAsnPheSerAspThrLysAsnTyrLeuProGluMetLysSerPhePhe 1120
 |||||
 Db 1023 CCTGACAGAAATTTAGTAAACACCAAAATTAACCTGCTCCAGAAATGAATTTTTC 1082
 QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
 |||||
 Db 1083 ACTCCTGGAAAACTTAACACCAATGTTCTCGGAGCAGT-AAATAGCCACTTTCATCA 1141
 QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
 |||||
 Db 1142 GCAGCAACAACTCTCAGACCAATCATCAAGAAATGAGAAATGTGAGCAACCAAGCAGAC 1201
 QY 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
 |||||
 Db 1202 AGCTCCAAACCCAGCTCTCTCGGAGAGTCAAGGAGGAGGCTTGATAGCTCTTAATGAT 1261
 QY 1181 HisSerGlnAsnGlnAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
 |||||
 Db 1263 CACATGTGAATAAGATTAATACATGTCTTCACTTTCCAGGAAAAAAGTGCACAG 1321
 QY 1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
 |||||
 Db 1322 AGAGAGAGACCCGTGATCTT-----TCTAGTGTGAGAGAACCTTAGAAGTCGAAAAAGCA 1375
 QY 1221 ProValThrGlnGlnGlnGluLysLeuGluMetAspAspLeuThrLysLeuValGlnGlu 1240
 |||||
 Db 1376 CCTGTCCAGACCCCTGAAGAGAAATTAAGTATGATGACCTAAGTAAAGTGTGTACAGGAA 1435
 QY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260

Db 1436 CAGAAACCTAAAGCAGCTAGCGAGGCCGGAAGAGGCCGTTACAGCTTACAGCTCAGAC 1495
 |||||
 QY 1261 GlnGlnGlnIleTyrProGluGlnLysArgLeuLysGlnAspIleLeuGlnAsnGlnAspGlu 1280
 |||||
 Db 1496 GAGCAGAGTGGCTGTGAGAGAGAGGACCAAAAGAGAGACTCTCTGAAAAATGAGATGAG 1555
 QY 1281 GlnAsnSerProProLysGlyLysArgGlyArgProProLysProLeuGlnLysGly 1300
 |||||
 Db 1556 CAGAACAGCCACCAAAAAAGGCAAAAGAGGAGGCCACCAAAACCTTGTGTGGGGGG 1615
 QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLysGlySerLysLysSerLys 1320
 |||||
 Db 1616 ACATGGAAGGAAGAGCCCAATGAAACATCCAGAAAGAAAGAAAAAACTTTGTA 1675
 QY 1321 ProProAlaProGluGlnGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
 |||||
 Db 1676 CCTCTGTAGTAGAGCAGCATGAAAGAAAGAAACAAATTTGAAACACAGACATTAAG 1735
 QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
 |||||
 Db 1736 TCAAAAGCAACAGACAGCCGACATCAAGAGACACACAGACAGAGCAATCTCTGAA 1795
 QY 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
 |||||
 Db 1796 ACAAGTGCAGTTGAATCCACACAGTCCACACACACAGAAAGAGAGAAAGACATCAAAA 1855
 QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
 |||||
 Db 1856 GCACCATCACATCAACACCCCAAAAAAATC 1888
 |||||
 RESULT 5
 BC032988
 LOCUS BC032988 1359 bp mRNA linear HTC 20-SEP-2002
 DEFINITION Mus musculus, clone IMAGE:1179696, mRNA.
 ACCESSION BC032988
 VERSION BC032988.1 GI:21426933
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1359)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
 COMMENT Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: Marcello Bento Soares, Ph.D.
 CDNA Library Preparation: Soares Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 Contact: amadanesystemsbiology.org
 Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
 Clome distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRK Plate: 66 Row: C Column: 13
 This clone has the following problem: no polyA-tail.
 Location/Qualifiers
 1..1359
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1179696"
 /tissue_type="Mammary gland"

FEATURES

source

```

/clone.lib="Soares_mammary_gland_NBMG"
/lab_host="DH10B"
/note="Vector: pT73D-Pac"
BASE COUNT      505 a      296 c      304 g      254 t
ORIGIN

```

Alignment Scores:

Pred. No.:	2,07e-153	Length:	1359
Score:	2170.50	Matches:	419
Percent Similarity:	95.148	Conservative:	12
Best Local Similarity:	92.498	Mismatches:	19
Query Match:	30.188	Indels:	3
DB:	11	Gaps:	2

QY Db QY

1259 S
957 T
1279 F

```

UserAspG1
|||||
CCAGACGA
AspG1uG1
|||||

```

OY	939	TyrMetAlaIleCysAlaLeuCysAlaAlaLysAspProValLysGluValGAGAlaHisAla	958
Db	3	TACATGGCGCATCTGTGCTCTTTGTGGCCAAAGACCTGTGAAAGAGAGCGAGCCATGCT	62
OY	959	ArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAla	978
Db	63	AGACAGTCTCTGGTGAAGAACATCATGCTGGAGGAGGAGTACCTGAAGCAGACATGACGCT	122
OY	979	ValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeu	998
Db	123	GTTAGTGAATAATTATGTCTCTCTTACACAGATGTGTTCATATACATTCACCTT	182
OY	999	LeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLys	1018
Db	183	TTGGCACATGACCCAGATTATGTCAAGATACAGATATTGACAACTTAAGATGTGAA	242
OY	1019	GluCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAla	1038
Db	243	GAAATGCTTGTGTTGTCTGGACATATGTGGCTATAAAATGAAACAAACAGCCATGCA	302
OY	1039	PheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAsp	1058
Db	303	TTTATTCAGAAAATAGTGAATAATTATTAACAGACAAAGATGCTCAAGACCGACATGAT	362
OY	1059	AlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSer	1078
Db	363	ACAAAATAAGATGAAAAATTTGTACACGCTGTGTATGTTGCCAAGACATCATCATGTCC	422
OY	1079	LysSerThrThrTyrSerLeuGlnSerProLysAspProValIleuProAlaArgPhePhe	1098
Db	423	AAGAGACACAGTACGCTGTGAGTCTCTAAGAGACCCGCTGTCGCCAGCTGGCTTTTTC	482
OY	1099	ThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSer	1118
Db	483	ACCAGCGCTGACAAGAATTTAGTAAACCCAAAATTAACCTGCTCCAGAAATGAATCA	542
OY	1119	PhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeu	1138
Db	543	TTTTTCACCTCCTGGAAAACTTAACACGCCATATCTTCGGAGGACGTTAATTAACCACTT	602
OY	1139	SerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAspAla	1158
Db	603	TCATCAGACGAGCAAAACAGTCTCAGACCAACATCATCAAGATGAACAACTGTGACACCA	662
OY	1159	SerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGlu	1178
Db	663	AGCAGACAGCTCCAAACCCCAAGCTCTCTTGAAGAGATCAAGAGGAGGCTTGATAGCTGAA	722
OY	1179	MetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSer	1198
Db	723	ATGATATCACTAGAAAATGAAAGATTATTAACAATGCTTTCACCTTGGCCAGGAAAAAAGT	782
OY	1199	AspLysAlaGaspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLys	1218
Db	783	GCACAGAGAGAGACCCCTGATCTT-----TCTGATGTTGGAGAGCCTAGAGATCGGAAA	836
OY	1219	LysThrProValThrGluGlnGluLysLeuLysLeuLysMetAspPheLeuThrLysLeuVal	1238

[illegible]

QY 932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaValAspProVal 951
 Db 1 CGCTTAGGGCTTCCCTTGGATGACATGGCCATCTGTGCTTTGGCCAAAGACCTGTG 60
 QY 952 LysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
 Db 61 AAAGAGAGCGAGCCATGCTAGACAGCTGTGGTGAAGAACATCAGTCTAGAGAGGAG 120
 QY 972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuProGluTyrVal 991
 Db 121 TACCTGAAGACGATGCGAGCTGTAGTGAATAATATTGTCTCTTCAACGAGATGTT 180
 QY 992 ValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIle 1011
 Db 181 GTTCATATACAAATTCACCTTTGGACATGACCAGATATGTCTCAAGTACAGATAT 240
 QY 1012 GluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLys 1031
 Db 241 GAACAACTTAAGATGTGAAGAAAGCCTTGGTTGGTTGTGAGATATGATGACTTAA 300
 QY 1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLys 1051
 Db 301 AATGAAACACACAGCATGATTTATCAGAAATAATGTAGAAATATTAACAGACAAAA 360
 QY 1052 AspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspVal 1071
 Db 361 GATCCTCAAGAGCAGATGATACAAATAATGATGAAATAATGTACACCGTGTGATGTT 420
 QY 1072 AlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspPro 1091
 Db 421 GCCATGAACATCATCATCTCCAGAGACGACGACGACGCTGGAGTCTCTTCAAGACCCC 480
 QY 1092 ValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyr 1111
 Db 481 GTGTGCGCAGCTCGGTTTTCACCCAGCCTGACAAAGATTTAGTAAACACCAAAATATAC 540
 QY 1112 LeuProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeu 1131
 Db 541 CTGGCTCCAGAAATGAATATCTTTTCTCTCTGGAATAAACAAGCCAAATGTTCTC 600
 QY 1132 GlyValAlaValAsnLysProLeuSerSerSerAlaGlyLysGlnSerGlnThrLysSerArg 1151
 Db 601 GGACAGATTATTAAGCCACTTTCATCAGCAGCAGCAACAGTCTCAGACCAANNATCATCAAGA 660
 QY 1152 MetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLys 1171
 Db 661 ATGGAAGCTGAGACAGCAAGCAGCAGCTCCANCCANCCCTGCTCGTGAAGATCAAG 720
 QY 1172 GlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSer 1191
 Db 721 GGGAGGCTTGTATGCTCTGAATGATCAGTCAAGTGAATAATGAATATATACATATGCTTCA 780
 QY 1192 ProLeu 1193
 Db 781 CCTTTG 786
 RESULT 8
 CA324284 779 bp mRNA linear EST 26-NOV-2002
 LOCUS CA324284
 DEFINITION UT-M-FY0-CCP-a-07-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
 IMAGE: 6822512 5', mRNA sequence.
 ACCESSION CA324284
 VERSION CA324284.1 GI:24542382
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 779)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLM at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 FEATURES
 source
 Location/Qualifiers
 1..779
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6822512"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1lb="NIH_BMAP_FY0"
 /note="Organ: Brain; Vector: pyx-Asc; Site:1: Ecor I;
 Site:2: Not I. The library was constructed according
 to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCAGACAG. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
 Developing Mouse Nervous System", supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator." 197 t 1 others
 BASE COUNT 241 a 166 c 174 g 197 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,95e-89 Length: 779
 Score: 1311.00 Matches: 255
 Percent Similarity: 98.46% Conservatve: 0
 Best Local Similarity: 98.46% Mismatches: 4
 Query Match: 18.23% Indels: 1
 DB: 14 Gaps: 0
 US-09-512-581b-2 (1-1391) x CA324284 (1-779)
 QY 863 AspleuthrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882
 Db 3 GATTGACAGAAACAGGAAATAATAGTAACACAGATATGTCACCGCTGAGACTGCTGCT 62
 QY 883 GlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGlu 902
 Db 63 GGGAGTGTATTGGAAGCTGGACAGAGCCCTGTTACACGAGATCATATTCAGTGAAG 122
 QY 903 GlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPhe 922
 Db 123 CAGTACAGCTGTGTGCTTATGACATCATATATAGTGTATTATCAAGTCAAGCAGGTGTC 182
 QY 923 AlaGlnLysLeuHisLysGlyLysSerArgLeuArgLeuProLeuGluTyrMetAlaIle 942
 Db 183 GCTCAGAAACCTTCAAAAGCGCTTCCCGCTTACGGCTCCCTTGAATACATGGCCATC 242
 QY 943 CysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
 Db 243 TGTGCTCTTTGTGCCAAAGACCTGTGAAAGAGAGGAGGCCATGCTAGACACTGCTTG 302
 QY 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIleValSerGluLys 982
 Db 303 GTGAAGAACATCACTGTGAGGAGGAGTACTGTAAGCAGCATWGCAGCTGTATGTAATAA 362

QY 983 LeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuAlaHisAsp 1002
 |||||||
 Db 363 TTAATTCCTCTTACACAGATATGTTTCATATACATTCACCTTTGGACATGAC 422
 QY 1003 ProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp 1022
 |||||||
 Db 423 CCAGATTATGTCAAACTACAGGATTTGAACAACCTTAAGATCTGAAGAATGCCCTTGG 482
 QY 1023 PheValLeuGluIleLeuMetAlaLysAsnGlnAsnSerHisAlaPheIleArgLys 1042
 |||||||
 Db 483 TTTGTTCTGAGATATTGATGCTAAATAATGAAACACGCCATTCATTTATCAGAAAA 542
 QY 1043 MetValGlnAsnIleLysGlnThrLysAspAlaGlnGlyProAspSpsAlaLysMetAsn 1062
 |||||||
 Db 543 ATGGTAGAAAAATTTAAACAGACAAAAGATGCTCAAGACCAATATACAAATGAAT 602
 QY 1063 GluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThr 1082
 |||||||
 Db 603 GAAAAATTGTACACCGCTGTGTGTGATGTTGCCATCAATCATCTATGCCAAGACACACAG 662
 QY 1083 TyrSerLeuGluSerProLysAspProValLeuProAlaArgPheThrGlnProAsp 1102
 |||||||
 Db 663 TACAGCCTGAGTCTCTTAAGACCCCGTCTGCCAGCTCGGTTTTCACCCAGCCTGAC 722
 QY 1103 LysAsnPheSerAspThrLysAsnTyrLeuProGluMetLysSerPhePheThr 1121
 |||||||
 Db 723 AAGAAATTTAGTACACACANAAATTAACCTGCTCCAGAAATGAA-TCATTTTTCACCT 778
 RESULT 9
 BU703488 786 bp mRNA linear EST 09-OCT-2002
 LOCUS UI-M-FOO-bco-g-20-0-UI.r1.NIH.BMAP_F00 Mus musculus cDNA clone
 DEFINITION IMAGE: 6405163 5', mRNA sequence.
 ACCESSION BU703488
 VERSION BU703488.1 GI:23629371
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 TITLES NIH-MGC (bases 1 to 786)
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 FEATURES
 source location/Qualifiers
 1..786
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6405163"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH-BMAP_F00"
 /note="Organ: Brain; Vector: pyx-Asc; Site1: EcoR I;
 Site2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TCAGACAGCC. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
 Developing Mouse Nervous System", supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

BASE COUNT 248 a 143 c 152 g 242 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 6.16e-88 Length: 786
 Score: 1298.50 Matches: 254
 Percent Similarity: 98.08% Conservative: 1
 Best Local Similarity: 97.69% Mismatches: 4
 Query Match: 18.05% Indels: 1
 DB: 13 Gaps: 1

US-09-512-581b-2 (1-1391) x BU703488 (1-786)

QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGluValLys 20
 |||||||
 Db 10 ATGGCTCATCAAGAA---AGACCAACGATGGGAAAAATTAATCTTACCTCTCGATCAAG 66
 QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValLys 40
 |||||||
 Db 67 GAAATCTCAGATTAATCTCTTAAGAGAGAGATGTGAGCGGTTAAAGATGCTTGAAAA 126
 QY 41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuLysLeuLysAla 60
 |||||||
 Db 127 ACTTTCATGACACTGGACCAAGACCTCGAAGAGAAAGAACTTATCTTAACCTAGCT 186
 QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
 |||||||
 Db 187 TTACATCTGCTTCTGCTGACTTCTCCCAAGACATCCGAAAGATGTTGTTTACCGTG 246
 QY 81 AlaCysCysLeuAlaAspIlePheArgGlyIleTyrAlaProGluAlaProTyrThrSerPro 100
 |||||||
 Db 247 GCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
 QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
 |||||||
 Db 307 GATAAACTAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 366
 QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
 |||||||
 Db 367 AAGAGCCCTCAATTTAATAGATATTTTATTTACTTGAGACATTCGATGGGTAAATCA 426
 QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
 |||||||
 Db 427 TATACATATGCTTGTGAGTTAGTAAGATACCAATGAATCTTCACTCACTTACAGAAC 486
 QY 161 LeuPheSerValIleLeuAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeu 180
 |||||||
 Db 487 TTAATCTCAGTTAATACATGAGCCCAATCAAGAAATGTCATATGACATGGTGACCTC 546
 QY 181 MetSerSerIleIleCysGluGluLysAspThrValSerGlnGluLeuLeuAspThrValLeu 200
 |||||||
 Db 547 ATGAGCTATATCATTTGTGACAGGTGATACCGTATCTCAGAGACCTGTACATACAGTTT 606
 QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrArgLeuAlaLysAla 220
 |||||||
 Db 607 GTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
 QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
 |||||||
 Db 667 TTGCTGAGAGAGACCTGCTAAGCTATGAAACCATATATTAACCAATTTTAAACCAAGGT 726
 QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
 |||||||

Db 727 CTGATGCTTGGAACATCTATGAGATTTTGTCTGAGACATGCTTGAATTGATCTGT 786

RESULT 10
 LOCUS BU363024 782 bp mRNA linear EST 28-NOV-2002
 DEFINITION 603790485F1 CSEQCHN72 Gallus gallus CDNA clone CHEST753121 5', mRNA sequence.

ACCESSION BU363024
 VERSION BU363024.1 GI:25871025
 KEYWORDS EST.
 ORGANISM Gallus gallus (chicken)
 SOURCE Gallus gallus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 782)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 2235534
 PUBMED 1243592

COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)

FEATURES
 source
)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..782
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHEST753121"
 /sex="Female"
 /tissue_type="cerebrum"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN72"
 /note="Organ: brain; Vector: pBluescript II KS(+); Site.1: EcoRI; Site.2: NotI. This normalized library was constructed from 1 million independent clones. CDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 279 a 146 c 160 g 197 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,12e-87 Length: 782
 Score: 1295.00 Matches: 253
 Percent Similarity: 98.85% Conservative: 5
 Best Local Similarity: 96.93% Mismatches: 2
 Query Match: 18.00% Indels: 2
 DB: 13 Gaps: 0

US-09-512-581B-2 (1-1391) x BU363024 (1-782)

QY 856 ThrlleuHhIsSerAspGlyAspleuThrluGlnGlyLysIleSerLysProAspMet 875
 |||||||
 3 ACAATTATTACACAGATGAGAGACTGTGACAGAGAGGGGAAATTAGTAAACCTGATATG 62

OY	876	SeRrIeUeUaRgLeUaLaIaIaGlySeRaIleValLySeUaIaGInGuIpRoCyS	Tyr	895
Db	63	TCTCGATTGAGGTTGGCCGAGGTAGTGCTATTGTGAAGCTGGCACAAGAACTTGT	TAC	122
OY	896	HisGluIleIeIeThrLeUeUgInUTrYcGInLeUcYsaLaLeuAlaIeasNaSPGIn	Cys	915
Db	123	CATGAATATCATCCTTAGAAGACATACCAGCTGTGTGCATAGCCATAAAGATGAATGT	T	182
OY	916	TyrTrIUnVaIaRgGInVaIaPheAlaGInLySLeUeHisLySGLYLeUSeRaTgLeUaRgLeU	Leu	935
Db	183	TACCAAGGAGGCAAAATTTTGGTCACAAACTTACAAAGGGCTTTCTTAACACTAGGGCTG	T	242
OY	936	ProLeUeUgIUrMeTaIaIeCySaLaLeUcYsaLaLySaSPProVaLySGLUaRg	Arg	955
Db	243	CCACTAGAAATATATGGCTATTGTGGCACTGTGTGCATAAAGATCCAGTGAAGAGAGAGA	G	302
OY	956	AlaHisAlaArGgInCYsLeUeValLySaSnILeasNaVaIaRgArGluTrYLeUySGLn	Gln	975
Db	303	GCTCATGTAGCGCAGTGCCTTGTGCAAAACATAAATCTCAGAAAGGAATATCTGAACAA	C	362
OY	976	HisLaIaIaVaISeRgInLySLeUeUSeRLeUeUProGluTrYVaIaIaProTrYThr	Thr	995
Db	363	CATCAGCAGCTTATGTAGAAACTTGTGCTTACCAGAGATATGTTGTTCCATTATACA	A	422
OY	996	IleHisLeUeUaLaHisAsPProAsPTrYVaLySaGInaSPILeGInLeUyS	lys	1015
Db	423	ATACATGCTTACGTCATGACCCAGATTAATGTCAAGGCCAGCATCGAAACAACATAAG	A	482
OY	1016	AspVaLySGLYcYsLeUTrPheValLeUgInUleUeUeTaIaLySaSnGuIaSaSn	Asn	1035
Db	483	GACATTTAAAGAGTGTGTGTGCTTACTCGAAATATATGATGGCTAAAAATGAGAACAC	A	542
OY	1036	SeRHisAlaPheIleArGlySMeTValGInuSnIleLySGLnThrLySaSPaIaGInGly	Gly	1055
Db	543	AGTCATGATTTATCAGAAAAATGTGTAAGAAATTTAAACAGACTAAAGATGCTCAAGA	G	602
OY	1056	ProAsPSPaIaLySMeTaSnGInLySLeUTrYThrVaLySaSPaIaIaMeTaSnIle	Ile	1075
Db	603	CCAGATGTGCAAAAAGAAAGAAAGAAACTCTACACACTGTGATGTAGCAATGAAATATC	A	662
OY	1076	IleLeUSeRLySSeRTrThr-TyrSeRLeUeUSeRProLySaSProVaIleUProAl	Ala	1095
Db	663	ATTATGTCAAGAGTACACACCATATCAGTTTGGAATCCCTTAAGATCTCTACTTCCAGC	C	722
OY	1095	aArGpHeThrGInProAsPlySaSnPheSeRaSnThrLySaSnTrYLeUProProGln	Gln	1115
Db	723	ACGATATGTACTCAACCTGCACAAGAA-TTTAGTAACACCAAAAATTAATCTCTCCAAA	A	781
OY	1115	u	1115	
Db	782	A	782	
RESULT 11				
LOCUS	B0914579	943 bp	mRNA	linear
DEFINITION	AGNCNCOURT.10496229 NICHED_XGC_OOI Xenopus laevis cDNA clone			
ACCESSION	B0914579			
VERSION	B0914579.1	GI:24096493		
KEYWORDS	EST.			
SOURCE	Xenopus laevis (African clawed frog)			
ORGANISM	Xenopus laevis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;			
TITLE	Xenopodinae; Xenopus.			
COMMENT	1 (bases 1 to 943)			
	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .			
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
JOURNAL	Unpublished			
COMMENT	Contact: Robert Strausberg, Ph.D.			
	Email: cga@bbs-1email.nih.gov			
	Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.			

Query Match:	88.61%	Mismatches:	21
Best Local Similarity:	17.65%	Indels:	2
DB:	14	Gaps:	1
US-09-512-581B-2 (1-1391) x CA987660 (1-857)			
QY 784	ProThrlySerSerTrpValAlaThrPheIleValIysAspLeuMetAsnAspArgLeu	803	
Db 3	CCCTTAATATCATGATGGTGGCCACCTCTTGTTAAAGATTGTGTATACAGATGGCTT	62	
QY 804	ProGlyIysIysThrThrIysLeuTrpValProAspGluIvalSerProGluThrMet	823	
Db 63	CCAGGTAAAAAGACACAAACAACTGGGTGCACATGATGAGGTGTCCAGCAAACTAG	122	
QY 824	ValIysIleGlnAlaIleIysMetMetValArgTrpLeuLeuGlyMetIlysAsnHis	843	
Db 123	GTTAAATTTCAAGCAATTAATAATGATGTGGGGTGGCTGGGAGATATAAATAATCTC	182	
QY 844	SerIysSerGlyThrSerThrLeuArgLeuThrThrIleLeuHisSerAspLysAsp	863	
Db 183	AGTAAATTCGGGAATTTCTACTTTACGACTCTTAATGGCAATATTGCACATGAGGAGAC	242	
QY 864	LeuThrGluGlnGlyIysIleSerIysProAspMetSerArgLeuArgLeuAlaIacty	883	
Db 243	TTTGACAGAACATGGAACCTAGTAACCAACACATGACGCGTAGACTGCTCTCC	302	
QY 884	SerAlaIleValIysLeuAlaGlnIuProCysTryHisGluIleIleThrLeuGluGln	903	
Db 303	AGTGCATTTGTTAACTAGCCGACGAGGCACTTACCACTGAAATATTACTTAGAACAG	362	
QY 904	TyrGluLeuCysAlaLeuAlaIleAsnAspCysTryGlnIvalArgGlnIvalPheAla	923	
Db 363	TATCAGCATGTGCTATGGTTCATCATATGATAGTGTCCAGGTTAGACAGTTGTTTGGCT	422	
QY 924	GlnIysLeuHisIysGlyLeuSerArgLeuArgLeuProLeuGluIuTrpMetAlaIleCys	943	
Db 423	CAAAAAATTCCAAAGGCGCTCCAGATTAAAGACTTCCATTAGATATATGCGCATCTGT	482	
QY 944	AlaLeuCysAlaIysAspProValIysGluArgAlaHisIalaArgGlnCysLeuVal	963	
Db 483	GCTGTATGTGCAAAAGACTCTGTGTAAGAGAGCGCGCTCACGCGACAGTGTGGTG	542	
QY 964	LysAsnIleAsnValArgArgGlnIuTrpLeuysGlnHisAlaIvalSerGluIysLeu	983	
Db 543	AAAAATATTATGTGCAGAGAGACTATTGTGAACGCAATGACGCTGATGAAAAATTG	602	
QY 984	LeuSerLeuLeuProGluIuTrpValIvalProTryThrIleHisLeuLeuAlaHisAspPro	1003	
Db 603	TTTCTCTTTCGGCCTGAATATGTGGTTCATATCTGTTCAATTGCTTGCAACATGACCA	662	
QY 1004	AspIlyrValIysValGlnAspIleGluGlnLeuIysAspValIysGluCys-LeuTrpPh	1023	
Db 663	GATTATGTGGAAGATACAGATATATGAGCGATTAAGGACATCAAAAGATGGCTGTGTT	722	
QY 1023	ValIleuGluIleLeuMetAlaIysAsnGlnIAsnAsnSerHisAlaPheIleArgIysMe	1043	
Db 723	CGTCTCGGAATTTCTTATGTCAAGAATGAATAATACAGCCATGCTTTATTCGAAAT	782	
QY 1043	IvalGluAsnIleIysGlnIuTrpIysAspAlaGlnGlyProAspAspAlaIysMetAsnG	1063	
Db 783	GGTGAATATATTATTAAGCCAAACCAAGATGAGCCCAAACTGATGACCGAGAG--AATGA	839	
QY 1063	u 1063		
Db 840	A 840		
RESULT 13			
LOCUS	B0916668	882 bp	mRNA linear EST 17-OCT-2002
DEFINITION	AGNOCOURT 10496784 NICHD_XGC_CO1 Xenopus laevis cDNA clone		
IMAGE	6643610 5', mRNA sequence.		
ACCESSION	B0916668		
VERSION	B0916668.1	GI:24098582	

KEYWORDS	EST
SOURCE	Xenopus laevis (African clawed frog)
ORGANISM	Xenopus laevis
REFERENCE	Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
AUTHORS	1 (bases 1 to 882)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.
COMMENT	Email: cgaps.rem@nhi.gov
COMMENT	Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
COMMENT	CDNA Library Preparation: Life Technologies, Inc.
COMMENT	CDNA Library Arrayed by: The I.M.A.G.E. Consortium
COMMENT	DNA Sequencing by: Agencourt Bioscience Corporation
COMMENT	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://lham14230.row.c
COMMENT	Plate: LHAM14230 row: 0 column: 02
FEATURES	High quality sequence stop: 683.
SOURCE	Location/Qualifiers
	1..882
	/organism="Xenopus laevis"
	/mol_type="mRNA"
	/db_xref="taxon:8355"
	/clone="IMAGE:6643610"
	/rlnsue_type="oocytes"
	/lab_host="DH10B (phage-resistant)"
	/clone_id="NICHD_XGC_001"
	/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Life Technologies."
BASE COUNT	310 a 169 c 192 g 210 t 1 others
ORIGIN	
Alignment Scores:	
Pred. No.:	3,69e-85
Score:	1262.50
Percent Similarity:	91.35%
Best local Similarity:	85.12%
Query Match:	17.55%
DB:	13 Gaps: 0
US-09-512-581B-2 (1-1391) x BU916668 (1-882)	
QY	956 AAlhAlaAlaAgGInCysLeuValLysAsnILleAsnValAArgAGLUTyrLeuLysGln 975
Db	3 GCTACGACGACGCGAGTGTCTGTGTGTAATAATTTATATGTCAGAAAGAGTATCTGAACAG 62
QY	976 HIsAlaAlaValSerGInLysLeuLeuSerLeuLeuProGInUTyrValValProTyrThr 995
Db	63 CATGCACTGTGAGTGAATAATTTGTCTCTCTCTGCTCGCTGAATATGTGGTTCATATACT 122
QY	996 ILleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspILleGInLLeuLys 1015
Db	123 GTTATTATTCGTCGACATGACCCAGATTATGTGAAGAGTCAAGATATAGACACTAAAG 182
QY	1016 AspValLysGInCysLeuTyrPheValLeuLInLleLeuMetAlaLysAsnGInAsn 1035
Db	183 GACATCAAAAGAGTCCCTGTGGTGTGTGTGTGAGATTCTTATGTCAAAGAAATGAATAATAC 242
QY	1036 SerHisAlaPheIleArgLysMetValGInAsnILleLysGInTyrLysAspAlaGInCly 1055
Db	243 AGCCATGCTTTTATTCGGAATAATGTGTGAATATTTTAAAGCAAAAGAAAGATGGCCAAAC 302
QY	1056 ProAspAspAlaLysMetLysGInLysLeuTyrThrValLysAspValAlaMetAsnIle 1075
Db	303 CCCGATGACCGAAGATGATGATGAATAATGTACACAGTGTGTGATGTGGCATGATATATA 362
QY	1076 ILleMetSerLysSerThrThrTyrSerLeuGInSerProLysAspProValLeuProAla 1095

Db 363 ATTATTTCAAGAGCAGCAACATACAGCTTAGAGTCCCGAGAGATCTGTATTGCTGCA 422
QY 1096 ArgPhePheThrGlnProAspLysAsnPheserAsnThrLysAsnTyrLeuProProGlu 1115
Db 423 CTTTCTTCACAGCAGCATAGAGATTTTCAGCAACAAAAAATATCTTCCCGCTGAA 482
QY 1116 MetLysSerPhePheThrProGluLysProLysThrAsnValLeuGluAlaValAsn 1135
Db 483 TTGAAGTCTCTTTTACACAGCAAGAAACCAAAATCTACTAACTTTTGGGTGAGTCAT 542
QY 1136 LysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal 1155
Db 543 AAACCTTATTCAGGCGCAGAGAAAGCAAAATGCTATCCAGTCATCTGATGAGACTGTA 602
QY 1156 SerAsnAlaSerSerSerSerAsnProSerSerProGlyArgLLeuLysGlyArgLeuAsp 1175
Db 603 ACCAATGCAAGCAGCGGTTCTTAACCCACCTCTCCAGGAGATAAAGAAACACTTGAC 662
QY 1176 SerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGly 1195
Db 663 ACTATGGAATTCGATCAAGTGAATGAACTATACATGCTCTCTCTCTCTCTCTCTG 722
QY 1196 LysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArg 1215
Db 723 AAAAAAAGTGAAGAAAGGAGCAGCTCGATCTCTGAACTGAGTTAGAAAAAGCCAAAG 782
QY 1216 GlyArgLysLysThrProValThrGlnGlnGlnLysLeuGlyLysMetAspLeuThr 1235
Db 783 CG-AGGGAACAAATCTCTATAGACCAAGATGACAGTTTAAAGTGAATTAATATC 841
QY 1236 LysLeuValGlnGlnLysProLys 1244
Db 842 AAACCAAGCAGGAGAAACCAATCAAG 868

RESULT 14
BU914025 931 bp mRNA linear EST 17-OCT-2002
LOCUS AGENCOURT.10494702 NICHD_XGC_001 Xenopus laevis cDNA clone
DEFINITION IMAGE:6640480 5', mRNA sequence.
ACCESSION BU914025
VERSION BU914025.1 GI:24095939
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Xenopus laevis
Amphibia; Batrachia; Chordata; Craniata; Vertebrata; Euteleostomi;
Xenopodinae; Xenopus.
1 (bases 1 to 931)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: NCI-CGAP Clone distribution Information can be
<http://image.llnl.gov>
Plate: L1AM1422 row: 1 column: 16
High quality sequence stop: 552.
Location/Qualifiers
1..931
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6640480"
/tissue_type="ooocytes"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC_001"
/note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI;

BASE COUNT 286 a 173 c 202 g 269 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 2,33e-83 Length: 931
Score: 1239.00 Matches: 253
Percent Similarity: 93.88% Conservative: 8
Best Local Similarity: 91.01% Mismatches: 14
Query Match: 17.23% Indels: 4
DB: 13 Gaps: 0
US-09-512-581b-2 (1-1391) x BU914025 (1-931)
QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGluValLys 20
Db 98 ATGGCTCTCCCAAGGCCCAAGGCAAGTGGAAAAATGACGATCCCGAGGTGCAAG 157
QY 21 GluIleSerAspLysIleSerLysGlnGluMetValArgArgLeuLysMetValLys 40
Db 158 GAATTTTCAGACAGATTTTCAGAGAAATGTTAGACGGTTAAAGATGGTTGTA 217
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
Db 218 ACCTTCATGATATGATGATCAGACCTCTGAAGAAAGAAAGAGACGATTTGAACCTTGGT 277
QY 61 LeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
Db 278 TTACACTTATGACCTCTACTCTTCTCCCAACATCTCGAACAAAGATGTAGATGCTTGTGA 337
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
Db 338 GCATGCTGCTGCTGATATATTTTGAATTTAGCCCAAGAGCTCATATACATGCGCT 397
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
Db 398 GATAAACCTGAGATATATTTATGTCATATACAGCAACMACTAAAGGCTCTTGAAGATACA 457
QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTyrValLysSer 140
Db 458 AAGAGTCCACAGATTATATGATATTTCTATTTGCTGAGAAATATTCATGAGGTAATCC 517
QY 141 TyrAsnIleCysPheGlnLeuGluAspSerAsnGlnIlePheThrGlnLeuTyrArgTyr 160
Db 518 TATATATATGTTTGTAGCTGGAGATGCAATGAGATTTTTCGACGCTATATAGAAC 577
QY 161 LeuPheSerValIleAsnAsnGlnHisAsnGlnLysValHisMetHisMetValAspLeu 180
Db 578 TTATTTTCGGTATCATATATGACATATATCAGAAAGTTCATATGATATGCGATTTTA 637
QY 181 MetSerSerIleIleCysGlnGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
Db 638 ATGAGCTATATGTTGTGAGAGAGATACAGTTTCTCAGAGAGCTCGATAGTGTCTTA 697
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLys 220
Db 698 GTGAACCTTGTCTTCACACAGAAATTTTAAACAACAAGCTTATACCTGCGAAAAAGCT 757
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
Db 758 CTGCTTAACAAAACCTCTACAGCTATTTGAACCATATATACAAA-TTTTTCACCAAGTT 816
QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisVal-PheAspLeu-Ileu 260
Db 817 CTCCTCTTGGCAAAACCTATATAGTATTTGTGAGAACCCCTGCTTGATGATGATTC 876
QY 260 euGlnLeuTyrAsnIleAspSerHis-LeuLeuLeuSerValLeuPro 275
Db 877 TGGAGCTTACACATTCGATACCATTTTGTCTTCTTCTGTTTAACT 924
RESULT 15
BX314916

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 25, 2003, 00:10:16 : Search time 852 Seconds

(without alignments)
4064.415 Million cell updates/sec

Title: US-09-512-581B-2

Sequence: 1 MAHSKRTNDGKITRYPGVK.....QKGRGRPSKTPSPQPKKNV 1391

Scoring table:

BLOSUM62			
Xgapop 10.0	Xgapext 0.5		
Ygapop 10.0	Ygapext 0.5		
Fgapop 6.0	Fgapext 7.0		
Delop 6.0	Delext 7.0		

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=x1h
-O=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163549_5590/app_query.fasta_1.1543
-DB=Published_Applications_NA -OFMT=fastlap -SUFFIX=trpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09512581.ecgn_1.1.474_errunat_23092003_163549_5590
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEOUTPUT -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-GAPOP=6 -GAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/1/pubpna/PCIT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/1/pubpna/PCITUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1409.5	19.6	2496	9	US-09-822-849A-496 Sequence 496, App

2	762.5	10.6	1217	9	US-09-925-302-357	Sequence 357, App
3	698	9.7	439	13	US-10-040-739-107	Sequence 107, App
4	598	8.3	471	11	US-09-918-995-3514	Sequence 3514, App
5	349	4.9	530	10	US-09-879-536-682	Sequence 682, App
6	276	3.8	433	11	US-09-918-995-20720	Sequence 20720, A
7	276	3.8	433	11	US-09-918-995-33817	Sequence 33817, A
8	232	3.2	8493	13	US-10-071-766-51	Sequence 51, Appl
9	231.5	3.2	9274	10	US-09-885-535-3	Sequence 3, Appl
10	221	3.2	8503	14	US-10-037-270-130	Sequence 130, App
11	229	3.2	7992	11	US-09-893-519A-140	Sequence 140, App
12	229	3.2	8491	12	US-10-133-013-360	Sequence 260, App
13	228	3.2	7034	14	US-10-118-513A-11	Sequence 11, Appl
14	225	3.1	5334	12	US-09-882-227-623	Sequence 623, App
15	205.5	2.9	3727	14	US-10-171-581-49	Sequence 49, Appl
16	205.5	2.9	7694	12	US-10-096-534-34	Sequence 34, Appl
17	205.5	2.9	8063	12	US-09-814-333-21776	Sequence 21776, A
18	205	2.8	7497	10	US-09-960-253-175	Sequence 175, App
19	205	2.8	7792	13	US-10-044-090-359	Sequence 359, App
20	200.5	2.8	4883	14	US-10-118-513A-5	Sequence 5, Appl
21	200.5	2.8	7215	14	US-10-118-513A-13	Sequence 13, Appl
22	200.5	2.8	7215	14	US-10-118-513A-15	Sequence 15, Appl
23	200.5	2.8	10281	14	US-10-084-817-361	Sequence 361, App
24	200	2.8	10281	12	US-10-240-965-201	Sequence 201, App
25	197.5	2.7	10300	10	US-09-960-253-145	Sequence 145, App
26	197	2.7	10190	10	US-09-864-864-292	Sequence 292, App
27	197	2.7	10211	10	US-09-954-456-1153	Sequence 1153, App
28	197	2.7	10211	10	US-09-967-768A-186	Sequence 186, App
29	197	2.7	10211	11	US-09-918-624B-2	Sequence 2, Appl
30	196	2.7	10096	10	US-09-960-253-163	Sequence 163, App
31	194	2.7	5967	14	US-10-171-581-158	Sequence 158, App
32	193	2.7	5612	14	US-10-128-714-1388	Sequence 1388, App
33	193	2.7	5612	14	US-10-128-714-6388	Sequence 6388, App
34	193	2.7	7611	14	US-10-128-714-388	Sequence 388, App
35	193	2.7	7612	14	US-10-128-714-3388	Sequence 3388, App
36	191	2.7	4455	11	US-09-998-027-7	Sequence 7, Appl
37	191	2.7	10302	10	US-09-782-378A-23	Sequence 23, Appl
38	191	2.7	16531	12	US-10-101-510-667	Sequence 667, App
39	190.5	2.6	2212	9	US-09-925-297-142	Sequence 142, App
40	189.5	2.6	1639	12	US-10-017-161-2225	Sequence 2225, App
41	189	2.6	5397	14	US-10-128-714-7388	Sequence 7388, App
42	189	2.6	12337	14	US-10-171-311-5	Sequence 5, Appl
43	189	2.6	12462	14	US-10-171-311-1	Sequence 1, Appl
44	187.5	2.6	6452	14	US-10-205-823-418	Sequence 418, App
45	186.5	2.6	5194	11	US-09-998-027-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-822-849A-496
Sequence 496, Application US/09822849A
Patent No. US20020045170A1
GENERAL INFORMATION:
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulakota, Kamalakkar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6403
CURRENT APPLICATION NUMBER: US/09/822,849A
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/195,582
NUMBER OF SEQ ID NOS: 598
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 496
LENGTH: 2496
TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-822-849A-496

Alignment Scores:
Pred. No.:      8,26e-118      Length:      2496
Score:          1409.50      Matches:      300
Percent Similarity: 65.96%      Conservative: 76
Best Local Similarity: 52.63%      Mismatches: 123
Query Match:      19.60%      Indels:      72
DB:              9          Gaps:      8

US-09-512-581b-2 (1-1391) x US-09-822-849A-496 (1-2496)

QY 717 LysSerLysGlyProPArgGlnAlaLysTyrAlaLeuHisCysIleHisAlaIle 736
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 4 AAGACAAAGAGGGGTACTCCACACACAGCAAAACAGGCTGCTCATATACAGCCATA 63

QY 737 PheSerSerLysGluThrGlnPheAlaGlnLeuPheGluProLeuHisLysSerLeuAsp 756
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 64 TTCACAAATTAAGAAAGTCCAGCTTGACACAGATTTTTCAGTCA----- 105

QY 757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuLeu 776
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 105 ----- 105

QY 777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp 796
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 105 ----- 105

QY 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlu 816
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 106 -----ACAGGTGAAAAGAAATGCAAAACGTGGTCTCCAGATGAA 144

QY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 145 GAGGTTTCCCTGAAAGTACTAGCAAGAGTTCAGGCATTAAACTTCTGGTAGGTGGCTG 204

QY 837 LeuGluMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThr 856
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 205 TTGGGTATGAAAAAACACACAGTCTAAATCTGCCAATCAACCCCTTGCTATTATTCAGCGC 264

QY 857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 265 ATGTGGGTATGAGGGTGTGAGCTGACACAGCAAAAGAGATCACTGAATATCTGATATCTCT 324

QY 877 ArgLeuArgLeuAlaIaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 325 CGCTGGGATTAGCTGCTGGTAGTGCATATATGAAGCTTGCTCGAAGACCTTGTTACCAT 384

QY 897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 385 GAATATTATTCACCCAGAACAGTTTCAAGCTGTGTCACCTGTATTATGATGAGATGTTCAC 444

QY 917 GluValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 445 CAAGTAAAGGACAAATTTGCTCAGAACTCATTAAGCGACTTGTGAAGTTAACTGCTCCCA 504

QY 937 LeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 505 TTGGAGATATTTGGCGATCTTGGCTTTGTGCGCAAGATCTCTGGAAGGAGAGAAAGCA 564

QY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 565 CACGACAGCAAAATGTTACTGAAATAATATCAGTATACGACGGAATATCAATTAAGCAAAAT 624

QY 977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 625 CCTATGGCTACTAGAAATATATATACCTGTGCTGAATAATAGTAGTTCATATCATATATT 684

QY 997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 685 CACCTGCTAGCCCATGATCCAGATTTTACAAAGATCACAAAGATGTGTATCAGCTTGAT 744
```

```
QY 1017 ValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsnGlnAsnSer 1036
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 745 ATCAAAAGATGCCATATGTTTCATGCTTGAAGTTTATTCACAAAGATGAAGAAACAATAGC 804

QY 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 805 CATGCCCTTTATGAAGAAGATGGCGAGAACATCAAGTTAACTCAAGAGATGCCAGTCTCCA 864

QY 1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 865 GATCAATCCAAAGACAAATGAATAAATGTATACATGATATGATGAGCTCTCTGTGTATATA 924

QY 1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 925 AATAGTAAAGTGGCTTTGT--GCAATGCAGATTTCCCAAGAGCCACCTCCCAATGAAT 982

QY 1097 PheThrThrGlnProAspLysAsnPhSerAsnThrLysAsnTyrLeuProProGluMet 1116
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 983 TTTTACACA---CCTGAAAAGGACTTCTGTAAACCATGAAGATTATATTTCAGAAAGACACA 1039

QY 1117 LysSerPheThrProGlyLysProLysThrThrAsnValLeuGluAlaValAsnLys 1136
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1040 AGACTACTTCTGTTAACAGGAAAGCCAAAGCCCTCTGAGTACTAGTGCAGTAAATTAAG 1099

QY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1100 CCTTATTCACACAGCAGAGGAAACCTTATGTATGAGACACCTGGCAGCTGAGACTGAAAGC 1159

QY 1157 -----AsnAlaSerSerSerSerAsnProSerSerProGluArgIleLysGlyArg--- 1173
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1160 AATATTATGTAAATTGATGAGCTGAACCTTCAACCGAAATGATCAAGAGAAACAGAGT 1219

QY 1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp-----TyrThrMetSerSer 1191
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1220 TCAGAGGCACAGAGAACTGGAGTATGTAAGAAATGAACAGAAACCTGTGAGAGATTATTTCA 1279

QY 1192 ProLeuProGlyLysLysSerAspLysArgAspSerSerAspLeuValArgSerGluLeu 1211
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1280 GTCAACACTGTAAAGATATATGACCAGTGAAGAAATGAG-----GTAATAATGCAATTGG 1333

QY 1212 GluLysProArgLysArgLysLysThrProValThrGluGlnGlu--GluLysLeuGlyMe 1231
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1334 CAAGGGAGAAATGAAGGCCAAACAGAACAGAGGCTCCAGCTTCTGCAAAAACCTTGATTT 1393

QY 1231 LAspAspLeu-----ThrLysLeuValGlnGlnGln-----Lys 1242
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1394 CACAATATGTCCTGTAACAGAAATGAAGCTCATTCAAGAACACACACTCTGCTTGAA 1453

QY 1242 sProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluI 1262
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1454 AACTAAAGACACTATCTTCTTTCACATGACACCAAGTCCCTGATGGAATGTACAA 1513

QY 1262 ngIntPrProGluGluLysArgLeuLys 1271
    |||:||||:|||||  |||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
Db 1514 GCAGAAACTCTTGAGAGAGAGGCTAATAA 1541

RESULT 2
US-09-925-302-357
; Sequence 357, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
```



```
; ORGANISM: Homo sapiens
US-09-879-536-682

Alignment Scores:
Pred. No.: 1.87e-22 Length: 530
Score: 349.00 Matches: 67
Percent Similarity: 98.53% Conservative: 0
Best Local Similarity: 98.53% Mismatches: 1
Query Match: 4.85% Indels: 0
DB: 10 Gaps: 0

US-09-512-581b-2 (1-1391) x US-09-879-536-682 (1-530)
OY 37 MetValValLysThrPhMetAspMetAspGlnAspSerGluGluLysGluLysGluTyr 56
DB 213 ATGGTGTGAAGACTTTATGATATGAGCCAGAGCTCTGAAGAAAGAGAGCTTTAT 154
OY 57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspVal 76
DB 153 TTAACCTAGCTTTACATCTTGCTTCAGATTTTTCCTCAGCATCTGATTAAGATGTT 94
OY 77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
DB 93 CGCTTACTGTAGAGCTGCTGCTGCTGCTGATATTTTCAGGATTTATGCTCGAAGCTCCT 34
OY 97 TyrThrSerProAspLysLeuLys 104
DB 33 TACACATCCCTCGATTAACCTAAG 10

RESULT 6
US-09-918-995-20720
; Sequence 20720, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20720
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-20720

Alignment Scores:
Pred. No.: 5.49e-16 Length: 413
Score: 276.00 Matches: 62
Percent Similarity: 52.94% Conservative: 10
Best Local Similarity: 45.59% Mismatches: 7
Query Match: 3.84% Indels: 57
DB: 11 Gaps: 1

US-09-512-581b-2 (1-1391) x US-09-918-995-20720 (1-413)
OY 11 GlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluLys 30
DB 7 GGCACGAGCGCTTACCTCCGCGGGGTAAAGATGATCAGACAGATCAACACGAGCAG 66
OY 31 MetValArgArgLeuLys----- 36
DB 67 ATGATCAAAAGCCGTGAAGGAGAAACAACAACTGCTTTTCACTTGAGCTGTCTAGAAAGGC 126
OY 36 ----- 36
DB 127 CCATGTAGTGTGGAATGAAAGCCTTTCAACAACAGCAAGTGTCAACTTATAACGTAGA 186
OY 36 ----- 36
```

```
DB 187 ATGACTTACCTTCAAAACACACCCCTCCAGCCCAAGCTTCAGATTAAGTACGAGCT 246
OY 37 -----MetValValLysThrPhMetAspMetAspGlnAspSerGluGluLysGlu 54
DB 247 TGACACAGATGTGAGTGAAGAACTTTATGATATGATGATCAGAGCTCAGAAAGATGA 306
OY 54 uLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLys 74
DB 307 GCAGATCTCCACACTAGAGCTTGACATCTGCAATTCCTTCCACAGAACCCCAATGA 366
OY 74 sAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
DB 367 AGATGTGCTCTCTCTTGTAGCATGTGTTGGCTGATATCTTTGCT 412

RESULT 7
US-09-918-995-33817
; Sequence 33817, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33817
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33817

Alignment Scores:
Pred. No.: 5.91e-16 Length: 433
Score: 276.00 Matches: 55
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.84% Indels: 0
DB: 11 Gaps: 0

US-09-512-581b-2 (1-1391) x US-09-918-995-33817 (1-433)
OY 1337 ThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAla 1356
DB 3 ACGGACAGAGAGTCCAAAGCAACAGCAGCAGAGTGTCAAGAGAGCAGCAGAGAGCA 62
OY 1357 GluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGly 1376
DB 63 GAATCTCTGATATGATGCAATGAAATGAAATGCAACAGCTCCACACACAGAAAGAGAG 122
OY 1377 ArgProSerLysThrProSerProSerLysProLysLysAsnVal 1391
DB 123 AGACCATCAAAAAGCCATCACATCACACCAACCAAAAAAATGTC 167

RESULT 8
US-10-071-766-51
; Sequence 51, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huel-Mel Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 51
; LENGTH: 8493
; TYPE: DNA
```


Db 2485 GGGAAACAAAAGATGACCTAGCACTACACAGTCGAATTATATAAGCACTGATCAAGAA 2544
 Oy 646 Ile----- 646
 Db 2545 TTCCAAATTTCCAAACCCCTTCATATGACCTTTAGCAAAAGCTATAGAGTGCCTTGAG 2604
 Oy 647 -----ArgAlaGlyLeuGluLeuLeuValLeuSerPheThrHisProIleSer 663
 Db 2605 GAGATGAGAGATGCAATCAGAAATAGCTATCTCTGAAA----- 2646
 Oy 664 PheHisSerAlaGlyThrPheGluSerLeuLeuAlaCysLeuLysMetasp----- 680
 Db 2647 -----GAAGCCCAAAATTTGATTCGAGTTTGCGTCTTGAAAGCCGAGCTTCTTAC 2700
 Oy 681 -----AspGluLysValAlaGluAlaLeuGluHisIlePheLysAsnThr 695
 Db 2701 AAGACCCAAAGACTTCAGAGAGAAACACGTGAGTTCCAAAGAAAGCTA-----AATGAG 2754
 Oy 696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
 Db 2755 ATGCAACAGCTGAGAAACAATTAGAAATAGAGATTCTCCGCTGCAAACTGATAGAAAG 2814
 Oy 716 HisLysSer----- 718
 Db 2815 GAGAAACACAGTATTACTGAGAAACTGACCAAACTTTAGAGAAGTAAAAACTTTAACT 2874
 Oy 719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
 Db 2875 CAGAAACAAAGATGATCTAAACAACTCCAAAGAAAGCTTGCAAAAT----- 2919
 Oy 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
 Db 2920 ---GAGAGGACCAAGCTCAAAAGATATGTCACGACTACTGTTAAATGATGAATGATACT 2976
 Oy 758 SerAsn-----LeuGluHisLeuIleThrProLeuValThrIleGluHis 772
 Db 2977 CAGAAACAATAGCAAAATGCTGTGATCTGCTGAAACAAACATCAAGAAACAATTAATACA 3036
 Oy 773 IleAlaLeuLeuAlaProAspGlnPheAlaIleProTyrLysSerTyrValAlaIleThrPhe 792
 Db 3037 CTA-----AAATCGAAATTTTCTGAGCA 3060
 Oy 793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysThrThrLysLeuTyr 812
 Db 3061 GTTTCACAGAAATTTCGATATG----- 3081
 Oy 813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
 Db 3082 -----GAGAAACAAATAGCAAGAACTTAAGATGAATTTTCAG----- 3117
 Oy 833 ValArgTyrPheLeuGluLysMetLysAsnHisSerLysSerGlyThrSerThrLeuArg 852
 Db 3118 ---CAAAAGATGGTGGCATGATTAATAAACAGATTTTGGAAGCTAAATAATCC---CA 3171
 Oy 853 LeuLeuThrThrIleLeuHisSerAspLysAspLeuThrGluGlnGlyLysIleSerLys 872
 Db 3172 ACACAACTGAGATGTTAAGATTAATAGAGTAATAGCAACAAGAAAGATATTTCT 3231
 Oy 873 -----ProAspMetSerArgLeuArgLeuAlaIleGlySerAlaIleValLysLeuAla 890
 Db 3232 TTAATACAGAGAAATAATGAACTCCCAAAATGTTAGAGAGTTTATAGCAAGAAAGCA 3291
 Oy 891 Gln-----GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
 Db 3292 CAATTTGAAGACTGACCTTAAGAGAAATATTTGAATGACCATTTGAAACAGCAAGAAATTA 3351
 Oy 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValAlaPheAlaGln 924
 Db 3352 AGACTT-----CTTGGGATGACTTAATAAAGCAACAAGATAGTTCACACA 3399
 Oy 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
 Db 3400 GAAAGAACCATGCGCATTAAGAAA-----GAAAGAGAGCTTTCTTAGAGCACTGTGAC 3450

Oy 945 LeuCysAla-----LysAspProValLysGluArgAlaHisAlaArgGlnCysLeu 962
 Db 3451 AGACTGGCAGAGATTGAGAGAAATAATTAAGAAAGCCCAAGCTCAAAAGAAACAG 3510
 Oy 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIleValSerGlu--- 981
 Db 3511 CAAACAATCTTAATGTACACAGACAGATGAGATGCAAGAAAAAGATTAATGAATTA 3570
 Oy 982 -----LysLeuLeuSerLeuLeuProGluTyrValProTyr 994
 Db 3571 GAGAAATTAAGAAATGAATTAAGAAAGAAAGAAATGAGATTTGCAATATGAAACAGAG 3630
 Oy 995 ThrIleHisLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
 Db 3631 AGCTTGATTTGGCTCAGAAACCTTAATGAAATTAATGAGAGATTAATCTTAACCAAA 3690
 Oy 1013 GlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsn 1032
 Db 3691 GAAAGAAAGTTCTAAAGGAA-----TTACAGAACTCATTTGAAACAGAG 3735
 Oy 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
 Db 3736 AGAGACCACTTAGAGATATATAGAGAAATTTGAGAGCTACAGGCTCAAAACCAAGAA 3795
 Oy 1053 AlaGlnGlyProAspAspAlaLysMetLysGlnLysLeuTyrThrValCysAspValAla 1072
 Db 3796 GAACTAATAATTTGCTCATATTTCACCTAAAGAAACCAAGAAACTTAATGATGATCAAGA 3855
 Oy 1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
 Db 3856 AGAACCGTA-----TCGAGAGAGCA----- 3876
 Oy 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnAsnPheSerAsnThrLysAsnTyrLeu 1112
 Db 3877 -----GCTCAAAATATTAATGACTCAGACCTTAGAAATAATCCATACCAAA-----TTA 3924
 Oy 1113 ProGluIleMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
 Db 3925 CAGAGAGATCCCA-----GTGCTTCAT 3948
 Oy 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser 1150
 Db 3949 GAGGAAACAAGATTAATCTGCTTAATGTAAGAAAGAACGACAGACCTCAGAGAAACAATGAT 4008
 Oy 1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
 Db 4009 GAACGTGAGATTATTAACAGACAGACGTCACCAACCAAGCACTCAACAACACTGCGCAAGAA 4068
 Oy 1171 LysGluArgLeuAspSerSerGluMetLysPheHisSerGlu---AsnGluAspTyrThrMet 1189
 Db 4069 -----GAAATGGAAGAGGCTCAGGTTGAATGAATAATTTCAAGAA 4107
 Oy 1190 SerSerProLeuProGlyLysLysSerAspLysArgAspSerSerLeuValArgSer 1209
 Db 4108 AGTCAGGAAGATTAATAATCTCTATACCAAGAAAGAACCAACCTTAATAACGATTAAGAA 4167
 Oy 1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu----- 1224
 Db 4168 GCCCTTGAAAGTTAAACATGACGACCTGAAAGAACATATTTAGAGAAACTTTGGCTAAATTC 4227
 Oy 1225 -----GlnGluGluLysLeuGlyMet-----AspAspLeu 1234
 Db 4228 CAGAGCTCAAAAGCAACAAGAAAGAGCTTAATTAATGAAGAAAGAAAGCAATGAAGTAACT 4287
 Oy 1235 ThrLysLeuValGlnGluGln-----LysProLys----- 1244
 Db 4288 ACAAATAATCGTAGAGATGAGCAATTCAAACCAAGAAATTCAGCACTAAGGATA 4347
 Oy 1245 -----GlySerGlnArgSerArgLysArgGlyHis----- 1254
 Db 4348 GAAATAGAAATGCTCGATGTGTCMAAAGACTTTCMAAGAAATGATGATGAATGAATCT 4407

Db 1423 TTGAAGCATCAGTAGATTTGGCTAAAGACTTGCACAGACCTTGTGTAGAGCCTCA 1482
 QY 361 AsprogluInuAlaIleargHisasprValIleValSerIleValThrlAlaIleAspLys 380
 Db 1483 GACTCAAAAGGGAACCC-----TCTGATGTTTTATTACAGTGAACCATAGCTCAAGAG 1536
 QY 381 AsplIleuLeuValAsnAsp----- 387
 Db 1537 CATAGTACACTAGCTACAGCAGCTTGATGAAAAAGTCTTCTTCTTAGAACCAAGCTTCAG 1596
 QY 388 -----HisIleuLeuAsnPhenValArgGlu----- 395
 Db 1597 GGCATTGGGCAATTCACAGAAATACCATTCGAGAAATGTTTCTCAGTTCGACAGAGTTGAT 1656
 QY 396 -----ArgThIleuAspLysArgTrp 402
 Db 1657 GATGAAGTGAATACAGCTGCTCCAGTGGGAGAGATGCGAAMAACCTTCGCAAAAGCAAAAG 1716
 QY 403 ArgValArgLys-----GluIleMetMetGlyLeuAlaGlnIleTrp 416
 Db 1717 GAAACTATAAAAGCCTTTCTAAAGAACTAGAACCCCTCATGTGCGAAGCAATGACATGCC 1776
 QY 417 LysLysTyralaIleuGlnSerAlaIleGlyLysAspAlaAlaLysGlnIleAlaTrpIle 436
 Db 1777 AATAAACCTCGAAGATGATGTTAGCCACAGAGAACCCTCTCGACCTGTGGGAATC 1836
 QY 437 LysAsp----- 438
 Db 1837 AAAAGGCACTTGAAGCCTTAAGCAAAACATGCAACAAGTTACTGGACCGCAAGCC 1896
 QY 439 -----LysIleuLeuHisIleTyTrpGlnAsnSerIle 449
 Db 1897 AGACAGACGAGCTTGAAGGACACATTAAGCCCTTGAAGAAATTTTACGCAAAATTTGAAA 1956
 QY 450 AspasprArgIleuLeuAlaGluArgIle-----PheAlaGlnTyMetValPro 465
 Db 1957 GAATTTCTATTCTGCTCCAGAAAGCCGAGAACATGAAAGTGCACAAAGTCCCTTGGT 2016
 QY 466 HisAsnIleuGluTrpThrlGluArgMetLysCysIleuTyTrpLeuTyAlaIleAsp 485
 Db 2017 ATGGAACGAGACAAATTAATACGACACCTTAACATGTTCAAGTATTCAGAAAAGAG 2076
 QY 486 Leu----- 486
 Db 2077 ATTGAACCTTCGAAGTAACACAGCAAGATGTAACCTGTAGTCAGAGCCTTATTTCAG 2136
 QY 487 AsnAlaValLysAla-----LeuAsnGlu 494
 Db 2137 AGTCTGTCGCAAAAGCACTAGCACTCAGGCTTGGAGCATGACCTGATGATGTCAATGCA 2196
 QY 495 MetTrpLys-----CysGlnAsnIleuLeu 502
 Db 2197 CGGTGGAAGACTCTCAATAGAGGTGGCTCAGCGAGAGCCGACGCTTCAGAGGCTTGG 2256
 QY 503 ArgHis-----GlnValLysAspIleuAspIleu----- 513
 Db 2257 CTGACACTGTGGAGATTCCAGATGCTCCCTGAGATCCCTGCTCAGCTGATGATGGACACT 2316
 QY 514 -----LysGlnProLysThrAspAlaSer-----ValLysAlaIle 525
 Db 2317 GAGAGAGCTTGGCCAAATCAAGAACCCCGTGGCTGAGTTCAAAAGTGTAAAGGCCAG 2376
 QY 526 PheSerLysValMetValIleThrArgAsnIleuProAspProGlyLysLysGlnAspPhe 545
 Db 2377 ATACAAAGCAAAAGCTTCTCCAGAGATGTTGGATGACCGAAAAATCTACGCTGAGAGTA 2436
 QY 546 MetLysLys-----PheThrGlnValLeuGluAspAspGluLysIle 559
 Db 2437 ATCAAGAGAGAGAGAAAAAATGCTACACACAGCAGCCCGCAGATAAAGTGAAGTT 2496
 QY 560 ArgLysGlnIleuGluValIleuValSer-----ProThrCys 571
 Db 2497 TTGAAACAGCTCACTCTTGTGATAGCAGATGGAGGCAATGCTTAATAAGCTGAACA 2556

QY 572 SerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLys 591
 Db 2557 AGGAATCGTCAAGTTGGAAAGT---ATCTCGGTGGAGCAGACGACATTTTCAGAAACCTTA 2613
 QY 592 GlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProVal 611
 Db 2614 GAACCACTGMAAGAGTGGCTTACAAACCATAGAAAAAGAGCTCGTGAATGTGAACCCATA 2673
 QY 612 HisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGly 631
 Db 2674 GGAAACCCAGCAATCTTAACCTTGAGGACAAACAAATTTGCGACAG---CACAAAGTTCTCGAAGAG 2730
 QY 632 ThrIleAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGly 651
 Db 2731 GACATCTTACTCGAGAAACAAATGTA-----GATCGGCTTTACTTAATAGTTAGAA 2784
 QY 652 LeuLeuLys-----ValLeuSerPheThrHisProIleSerPheHis 665
 Db 2785 CTACTTAACAAACCAACGAGTGCATCACTTTTAATTAATCAAGATTAATGGAAGCCATT 2844
 QY 666 SerAlaGluThrPheGluSerIleuAlaCysIleuLysMetAspAspGluLysValAla 685
 Db 2845 AAAGCAAGGTACAAAGAC-----ATTACTAACTGAGCACTGATGTCGCAAGACTCTGG 2898
 QY 686 GluAlaAlaIleuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHis 705
 Db 2899 GACAGGCGCTGCACCTTGCAGAGCGGCTGCAGCTCCACACAGAAAGACTGTGTAACCTGG 2958
 QY 706 IleArgSerAlaLeuLeuProValIleuHisLysSerLysGlyProProArgGln 725
 Db 2959 CTGACCAAGAGTGGAGTGAATTTACTTTCATATGAACCTCAG----- 3000
 QY 726 AlaLysTyralaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAla 745
 Db 3001 -----CTTCGAAGGAGAAAGCAAGTCAAGCA 3030
 QY 746 GlnIlePheGluProLeuHisLysSerLeuAspProSerGlnIleuGlnIleuIle-Th 765
 Db 3031 CAATG---ACACCAAGAGAACTGMAAAGAGAACTGAAGAACAAACAAACCTTACTGGAC 3087
 QY 765 rProLeuValThrIleGlyHisIle----- 773
 Db 3088 TCCCTTAATGAAGTGAAGAGAGCTTTCGTGTAAGCTGATGAGGCGCAAGAGAGCA 3147
 QY 774 -----AlaLeuLeuAlaProAspGlnPheAlaIlePro----- 784
 Db 3148 CTGGAAGAAATGTAGTGAAGACAAATGAGCGCTAACGATTAAGTGAAGCGACATGCACT 3207
 QY 785 -----TrpLysSerTrpValAlaIleThrPhe-----IleValLysAspPhe 797
 Db 3208 CAGAAAGTGGAGAGAGATGAGTGAAGCCATCTCTGCGATCAGACAGCTTGGACCAAGCACT 3267
 QY 797 uLeuMetAsnAspArgIleuProGly-----LysLysThrThrLysLeuTrpVal 813
 Db 3268 GATGCTGAGT-----TATCTGATGATTGCAAAACAGAAAAAATGATGCTGGGT 3321
 QY 813 lProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetVal 833
 Db 3322 GACATCAGAGCTT-GAGCAAGACCAAGACTTCTGCTCAGCTTCAAGTTCAAAG----- 3372
 QY 833 lArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerLysIleThrIleuArgGly 853
 Db 3373 -----ACATTCACCAATGAGAT 3389
 QY 853 uLeuThrThrIleLeuHisSerAsp-----GlyAspLeuThrGlnGluGly---LysIle 870
 Db 3390 TTTGAGA-----CACAAAGATATTGATGATGCTTGTAAATCTGGCATAAAT 3440
 QY 870 e-----SerLysProAspMetSerArg----- 877
 Db 3441 CATGACGCGATGCAAGTGAAGAGAAAGCAATCAATGAGAAAAAACTGCAAGGTAAT 3500

```

QY      878 -----LeuArgLeuAla1 883
Db      3501 GAAGAACTATGATACACTGCGAGATTAAATTCAGAAAGTATGTCAGCTGGAACGGC 3560
QY      883 ySerAla1leValLysLeuAlaIngluProCysThyHISgluLe-----Ileth 900
Db      3561 ACAGTCCCTGGTTAAACCAATTCGGAA---ACATATGAAGAACTTTGGCCATGCGTAC 3617
QY      900 rLeuGluIn-----TyrGluLeuCysAlaLeuAla1le----- 911
Db      3618 AGAAACACATCATCATCTCTGACGCTCCGCCGCCACCTTGAATATGAACTCTAAG 3677
QY      912 -----AsnAspGluCysTyrGluInValArgInValPheAlaGln----- 924
Db      3678 GCACGACGACGAGAACATCGGCAACTGCGTAGTTGATAGCTGAACACAAAGCCTCATAT 3737
QY      925 -----LysLeuHISLys---GlyLeuSerArgLeuArgLeu---ProLeuGlu----- 938
Db      3738 AGATAGATGACAAACCTGGCCACAGTTACTGGAATGAGCCCTGGGGAAGGCTTTTC 3797
QY      939 -----TyrMetAla1leCysAlaLeuCysAla-----LysAspProVally 952
Db      3798 TATCCAGAGAGATGATGCGACGCCGACACCTTTACAGTCAATTAAGAAAGATGCA 3857
QY      952 sGluArgArgAlaHISAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTy 972
Db      3858 AAAGCCGCTGTG-----GCACGTGATGAAGCCATTTCTCATCACTCACTGATT 3905
QY      972 rLeuLysGlnHISAlaValSerGluLysLeuSerLeuLeuProGluTyValva 992
Db      3906 CCATGACACAGATAGATAGATCCCTTGAGACCTGGAACGCCATCTGGACGCTGAGGCA 3965
QY      992 lProTyThrIleHISLeuAlaHISAspProAspTyrValLysValGln----- 1009
Db      3966 GCCACCCCTGATC-----TCTGCAGAGGTGAGAAAGATCAAGAACAGATCACTGA 4016
QY      1010 -----AspIleGluInLeuLysAspValLysLysLysLeuTyTrpPh 1023
Db      4017 AAATAGAAATGTGTCACTAGCATGAGAAAGCTACAGCCGTTGTATGAACCTTTAAACA 4076
QY      1023 eValLeuGluIleLeuMetAlaLys----- 1031
Db      4077 GAGGGGAGAGAAATGTTGCTAGATCTGGGGGACATGAAGACATATCTGCCAAGC 4136
QY      1032 -----AsnGluAsnSerHISAlaPheIleArgLysMetValGluAsnIleLysG1 1049
Db      4137 TGTTCAGATTAAGCTGACCAAAATGTTTCATTTGGGAGAACATACACACACTGNGGA 4196
QY      1049 nThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyThrValCy 1069
Db      4197 AGAGAGGGAAGCCAAACTACTGTGATGAGAGCTAGCAGAAAGTTCTGG-----TG 4250
QY      1069 sAspValAlaMetAsnIleIleMetSerLysSerThrThrTySerLeuGluSerProLy 1089
Db      4251 TGATCACATGTCATGTATAGTTACATTAAGATACATCAAGATTCAATCCGGGACCTGGA 4310
QY      1089 sAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnHeserAsnThrLy 1109
Db      4311 AGATCCCGAATGATTCCTTCAGTAAACAAACACACAGAACACAGACACATTAAG 4370
QY      1109 s-----AsnTyLeuProProGluMetLysSerPhePheThr----- 1121
Db      4371 GGAAGAAATAGATGACTACAGAGAGCTGATATATTAACCTAGATTCTGAAC 4430
QY      1122 -----ProGlyLysProLysThrThrAsnValLeuGlyAlaValas 1135
Db      4431 CATTCGGCATGTGGGAGCTGATTAACCATTTGTCAAGAAAGTATAGTGAATTAA 4490
QY      1135 nLysProLeuSerSerIaGlyGlnSerGlnThrLysSerSerArgMetGluThrva 1155
Db      4491 TTCAGCATGGGATTCCTTAATAAAGCTTGAAAGACCGGATTCACAAACTTGAGGAGCC 4550
QY      1155 lSerAsnAlaSer----- 1159

```

```

Db      4551 AATGACGCGTCCGTTCACTAGTACCAGATGAGCTGACAGCGGATTTTACTGGCTAGATAT 4610
QY      1160 -----SerSerAsnProSerSerProGlyArgIleLysGlyArgLe 1174
Db      4611 TGCAGGTGTAATTATTCCTTCAATGTCCTCAATTTGGAACAGATCTGAAACTGTCAAGCA 4670
QY      1174 uAspSerSerGluMetAspHisSerGluAsnGluAspTyThrMet----- 1189
Db      4671 GCAGATTGAAGACTTAAGCAATTTAAGTCTGAGGCCCTATCAACAGACAGATAGAAATGCA 4730
QY      1190 -----SerSerProLeuProGlyLysLysSerAspLysArgAspSerAs 1205
Db      4731 AAGACTGAATCATCAACAGACAGGCTTTGCTAAAGAAAGTAAACAGAGAGAGTCAACAA 4790
QY      1205 pLeuValArgSer-----GluLeuGluLysProArgGlyArgLysLysThrProva 1222
Db      4791 CACTGTTCAAGACCCATTAAATGGAACATGAAATTTGATATGGGATAGCCTGAGAGAGAA 4850
QY      1222 lThrGluGlnGluLysLeu-----G1 1230
Db      4851 CATCAACAGACAGCATTAACCTGGAGGGTGTCTATTAGCCTGGGTCAGTTCCAAATGC 4910
QY      1230 yMetAspAspLeuThrLys-----LeuValGlnGluGlnLysPr 1243
Db      4911 CCTGGATGACCTCTCTGCAATGCTTGACACACACCGAGGCTTCTAAGTGAAGAAACC 4970
QY      1243 oLysGlySer----- 1246
Db      4971 TGTTTGGAGGAGACCCCTTAACCATTTGAATTTGAACCTTGCACACATCATGCTCCAAA 5030
QY      1247 -----GlnArgSerArgLysArgGlyHISThrAlaSe 1257
Db      5031 TGATGTATTAAGCCATGATCAGTCCACAGTGAAGCGTTAATAAAGACAGAAATGATCTAAT 5090
QY      1257 rGluSerAsp-----GluG1 1262
Db      5091 TGAATCAAGTGCACGAGAGAGACAGCAACCTTCACAGACACAGCTAGAGGTTTAAATCA 5150
QY      1262 nGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluInas 1282
Db      5151 ACGCTGG-----CAAAATGTTTGGAAAAAAGACAAACAAAGGAA 5189
QY      1282 nSerPro-----ProLysLysGlyLysArgGlyArg----- 1292
Db      5190 GCACGACGTGATGTGCTCTTGCGCCAGGCCAAAGGGTTCATCAAAACCGCTGGAGGTTT 5249
QY      1293 -----ProProLysProLeuGlyGly-- 1299
Db      5250 GCAGCACTGGCTGACTGACACAGGAGCGTCATCTGTTGGCATCAAAACCGCTGGAGGTTT 5309
QY      1300 -----GlyThrProLysGlu----- 1304
Db      5310 ACCGGAAACGCCCAAGGAGCAAGCTTAATGTCCATATGAAGTCTGTGCTCCCTTTGAAGC 5369
QY      1305 -----GluProThrMetLysThr---SerLysLysGlySerLysLysSerGlyProp 1322
Db      5370 TAAAGAAAGAAATATTAAGAGTGTGATGAGCAAGAAAGCGACAGATGCTTCAAGATGCC 5429
QY      1322 oAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGlnGluSerLys 1342
Db      5430 A---AAATTCGACAGAGCAAAATATTGACCAAGACATAAATTAACCTTGGAAGAAATGGA 5486
QY      1342 sSerLysGlnHISArgValSerArgArgAlaGlnArgAlaGluSer 1358
Db      5487 ATCGGTGGAAACCAACTCAATGAAGAAACTAAATCTGGAAGAGGCT 5535

```

RESULT 10
 US-10-037-270-130
 ; Sequence 130, Application US/10037270
 ; Publication No. US20030104529A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom

```

/ APPLICANT: Liu, Chenghua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Yang, Yonghong
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yuning
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Tillinghast, John
/ APPLICANT: Dmanec, Radoje T.
/ TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
/ FILE REFERENCE: Polypeptides
/ CURRENT APPLICATION NUMBER: US/10/037,270
/ PRIOR FILING DATE: 2002-01-04
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 1104
/ SOFTWARE: PL-FL-genes Version 1.0
/ SEQ ID NO 130
/ LENGTH: 8503
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (91)..(8082)
/ US-10-037-270-130

```

```

Alignment Scores:
Pred. No.: 7,36e-10 Length: 8503
Score: 231.00 Matches: 285
Percent Similarity: 35.41% Conservative: 267
Best Local Similarity: 18.28% Mismatches: 541
Query Match: 3.21% Indels: 467
DB: 14 Gaps: 67

```

US-09-512-581b-2 (1-1391) x US-10-037-270-130 (1-8503)

```

QY 17 ProGluValLysGluIleSerAspLysIleSerIysGluGluMetValArgArgLeuLys 36
DB 1087 CCTATGCTTAAATGAGGATCA-----ACTGATGAAGCTCTCTGAAAAGGTATAGA 1137
QY 37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu----- 51
DB 1138 -----AAAGAAATATGATGATCTTAAAAACAATTAAGAGAGGTTTCTTTAGAGACG 1188
QY 52 -----GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
DB 1189 CGGGCTCAGCAATGGAAGAAAGCAATTTGGCCCACTTTTGGAAAGAA-----AAAGAT 1242
QY 67 PhePheLeuLysHisProGluLysAspVal-----ArgLeuLeuValAlaCys 82
DB 1243 TTGGCTTGAGAAAGTACAGAAATGAGAAATTTGAAAACCTTAACACGCGATCTGGTG----- 1296
QY 83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
DB 1297 -----ACCTCTTTCTTCC----- 1308
QY 103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuGluAspThrLysSer 122
DB 1309 -----CTCAGCTTTCGCA-----CAGGAATTAAGAGCT 1335
QY 123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaThrValLysSerTyrAsn 142
DB 1336 AAAAGAAAACGAGATTACTTGGTCTTGCAAAATTAACAAATGAAGAACTCAAC 1395

```

```

QY 143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrIlePhe 162
DB 1396 -----TATGCAAGATCAATTT 1410
QY 163 SerValIleAsnAsnGluHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
DB 1411 AATATACCAACAAT---ATAACACAAACAAACACATTAAGCTTTCTATGAATTAATTATACGA 1467
QY 183 SerIle-----IleCysGluGluLysAspThrValSerGlnGluLeuAspThrVal 199
DB 1468 GAATATGATGATCTGCTGTTCAGAGCTGATGTTTTCAGTAACACACTTGATACATTA 1527
QY 200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
DB 1528 AGTGAGATGAAATGGAATTCACCAACCAACCTACTAAATGACGAGAAATATAGAA----- 1581
QY 219 LysAlaLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrPhePheAsn 238
DB 1581 ----- 1581
QY 239 GlnValLeuMetLeuGluLysThrSerIleSerAspLeuSerGlnHisValPheAspLeu 258
DB 1582 -----AGTACGTTGAACCTCACTTCTGCTGACTATGATATCTCG 1620
QY 259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuLys 278
DB 1621 GTATTA-----GACTATGACAACTACACACAGCAAGAAAAAAATATGGAA 1665
QY 279 PheLysLeuLysSerAsnAspAsnGlnGluArgLeuGlnValValLysLeuLeuAlaLys 298
DB 1666 TTGAATTTAAAGAAAGAAATGATTTGGATGATGATTTGAGCGCTCTGAAAGAAACCTAAA 1725
QY 299 MetPheGluAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuThrGlnCysTyr 318
DB 1726 -----AAAGATCAAGACATGCCAA----- 1743
QY 319 LeuGluArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
DB 1743 ----- 1743
QY 339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
DB 1744 -----CTAATTCATGAAATTTTTCGAACTTAAGAAATTTAGTAAAGCATCGAAGTATAT 1797
QY 359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
DB 1798 AATCAAGATCTTGGAGAAATGAACTC-----AGTTCA 1827
QY 379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
DB 1828 AAAGTAGAGCTGCTTAGCAAAAGCAACACAGATTAAGAAAGCTACAGAAATACATACAC 1887
QY 399 AspLysArgTyrArgValArgLysGluValMetMetGluLeuAlaGlnIleTyrLysLys 418
DB 1888 TCTCAAAAG-----CTAGAAATATTAATAATGAGCTTGCA----- 1923
QY 419 TyrAlaLeuGlnSerAlaIleGluLysAspAlaAlaLysGlnIleAlaThrIleLysAsp 438
DB 1924 TACCTCATGGAAGAGATGGAAGACCCCAAAACAAATGAAGCAGACTCTTGTGATGCTGAA 1983
QY 439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuGluAlaArgIle 458
DB 1984 ACTGTA-----GCCCTTGATGCCAAGAGA 2007
QY 459 PheAlaGlnTyrMetValProHisAsnLeuGluThrGlnAlaTyrMetLysCysLeuTyr 478
DB 2008 GAATCAGCGCTTTCTTAGAAGTGAATCTGGATGGAGGAAGAAATGAAGAACTTGCA 2067
QY 479 TyrLeuTyrAlaThrIleAspLeuAsnAlaValLysAlaLeuAsnGluMetThrLysCys 498
DB 2068 ACTACATACCAAGCAAAATGGAA-----AATGATATTCAGCTTATAT 2106

```

QY 499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProLysThr 518
 Db 2107 CAAGCCAAATGGAGCGCAAAAGAAATGCAAGCTGATCTCGAAGAAAGAA 2157
 QY 519 AspAlaSerValLysAlaIlePheSerLysValMetValIlePheArg 534
 Db 2158 -----TTACATCTGCTTTTAATGAGATTAACAACAACTCCCTCCCTTATGAGTGC 2208
 QY 535 -----AsnLeuProAspProGlyLysIleGlnAspPheMet 546
 Db 2209 AAAGTCCAAAGATTTGCTGCTGTAATTTGGAAATTTGAGAGAAAGATTTACTGATCTTCAG 2268
 QY 547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
 Db 2269 AAAGAACTAATTAAGAAGTTGAAGAAATGAAAGCTTTGCGGGAAGAAAGTC---ATTGTC 2325
 QY 567 ValSerProThrCysSerCysLysGlnAlaGluLysValArgGluIleThrLysLys 586
 Db 2326 CTTTCA-----GAATTGAAATCTTTACCTTCTCAAGTAGAAAGCGCTGAGAAAGAG 2376
 QY 587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
 Db 2377 ATA-----CAAGCAAAATCTGAGAGAGCTCCATATATACATCAGAAAGAAAGAT 2424
 QY 607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
 Db 2425 AAATTTGTTTCTGAAAGTAGTTCATATAGAGAGTAGAGTTCAAGGTTTACTTGAAGAAAT 2484
 QY 626 AsnLysSerIleAspGlyThrAlaAspAspGluAspLysValAlaProThrAspGlnAla 645
 Db 2485 GGGAAACAAAGATGACTAGCAACTACACAGTCGAATTATTAAGACCTGATCAGAA 2544
 QY 646 Ile----- 646
 Db 2545 TTCCAAATTTCAAACCTTCATATGAGCTTTGAGCAAAAGTATAGATGCTGCTGAG 2604
 QY 647 -----ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
 Db 2605 GAGATGTAGAGATGAATCAGAGAAATAGTTAATCTCTCTAA----- 2646
 QY 664 PheHisSerIleGluThrPheGluSerLeuLeuAlaGlyLeuLysMetAsp----- 680
 Db 2647 -----GAAGCCCAAAATTTGATTCGAGTTGGGCTTTGAAAGCCGAGCTTTCTTAC 2700
 QY 681 -----AspGluLysValAlaGlnAlaAlaLeuGlnIlePheLysAsnThr 695
 Db 2701 AAGACCCAAAGACTTCGAGAGAAACGCGAGGTTCAAGAAAGACTA-----AATGAG 2754
 QY 696 GlySerLysIleGluGluAspPheProHisIleThrSerAlaLeuLeuProValLeuHis 715
 Db 2755 ATGAGACAGCTGAAGAAACATTAAGAAATAGAGATTCTCCGCTGCAAACTGTAGAAAG 2814
 QY 716 HisLysSer----- 718
 Db 2815 GAGAAACACTGATTTACTGAGAAACCTGACCAAACTTTAGACAGAAAGTAAAACTTTTACT 2874
 QY 719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
 Db 2875 CAGAAAGAAATGATGTAAACAAACACTCCAGAAAGCTTGCAAAAT----- 2919
 QY 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
 Db 2920 ---GAGAGGAGCAACTCAAAAGATGATATTCAGGATCTGTAAACATGATGATGATACT 2976
 QY 758 SerAsn-----LeuGluHisLeuIleThrProLeuValIleThrIleGlyHis 772
 Db 2977 CAAGACAAATTTAGCAAAATGCTTGAAGTCTGTAAGACACATCAGAAACAAATTAATACA 3036
 QY 773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerThrValAlaThrPhe 792
 Db 3037 CTA-----AAATCGAAAAATTTCTGAGAA 3060
 QY 793 IleValLysAspLeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuThr 812

Db 3061 GTTCCCGGGAATTTGCATATAG----- 3081
 QY 813 ValProAspGluValLeuSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
 Db 3082 -----GAGGAAATATACAGAGAAACCTAAAGATGAAATTTGAG----- 3117
 QY 833 ValArgThrProLeuGluMetLysAsnAsnHisSerLysSerGlyThrThrSerThrLeuArg 852
 Db 3118 ---CAAAAGATGGTTGGCATATAGATTAATAAACAGAGATTTGGACGCTAAAAATACC---CA 3171
 QY 853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
 Db 3172 ACACTAAGCTCAGATGTAAAGATTAAGATATATGACGACAAAGAAAGAAATTTTCT 3231
 QY 873 ---ProAspMetSerArgLysArgLeuAlaAlaGluLysSerAlaIleValLysLeuAla 890
 Db 3232 TTAAATACAGAGAAATAAGACTCCAAACAAATGTTAGAGAGTGTATAGCAGAAAGAA 3291
 QY 891 Gln-----GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
 Db 3292 CAATTTGAAGACTGACTAAAGAAATATTTGAATGACCATTTGAAACCCAGAAAGATTA 3351
 QY 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
 Db 3352 AGACTT-----CTTGGGAGTAGAACTTAAAGACACAGAGATGTTCGACAA 3399
 QY 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluLysThrAlaIleCysAla 944
 Db 3400 GAAAGACCAATGCCATTAAGAA-----GAGAGAGAGCTTTCTTACGACCTGTGAC 3450
 QY 945 LeuCysAla-----LysAspProValLysGluArgAlaHisAlaArgGlnCysLeu 962
 Db 3451 AGACTGGCAGAAAGTTGAAGAAACCTAAAGAAAGCCAGCACTCAGAAAGAAACAG 3510
 QY 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
 Db 3511 CAACAACTCTTATATGTACAAAGAAAGATGAGTGCAGTGCAGAAAGAAATTAATGAATA 3570
 QY 982 -----LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
 Db 3571 GAGAAATTAAGAAATTAAGAAACAAAGAAATTTGACATTTGAGACCTATAGAAACAGAG 3630
 QY 995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
 Db 3631 AGGCTTGATGGTGCAGAAACCTTAATGAAGAAATTAATGAGAAAGTATTAACCAAA 3690
 QY 1013 GlnLeuLysAspValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsn 1032
 Db 3691 GAAAGAAAGCTTCTAAAGGAA-----TTACAGAAAGTCAATTTGAAACAGAG 3735
 QY 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
 Db 3736 AGAGACCACTTAGAGATATATAGAGAAATTTGAAAGCTCAGAGCTCAGAAACCAAGAA 3795
 QY 1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
 Db 3796 GAACATAAAATGCTCATATTTCACTTAAGAAACCAACCAAGAACTATGATGAACATAAGA 3855
 QY 1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
 Db 3856 AGAAGCCTA-----CTTGAAGAGACA----- 3876
 QY 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnProPheSerAsnThrLysAsnTyrLeu 1112
 Db 3877 -----GCTCAAAATATATATACTCAGAGACTTTGAAAAATCCATACCAAA-----TTA 3924
 QY 1113 ProProGluMetLysSerPhePheThrProGluLysProLysThrThrAsnValLeuGly 1132
 Db 3925 CAAGAGAGATGCCA-----GTGCTTCAT 3948
 QY 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser 1150

```

Db      3949 GAGGAACAAGATTACGCTTAATGTGAAAAAAGTCAGTACGAGCTCAGAGAAACATGAAAT 4008
QY      1151 ArgmetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
Db      4009 GAACGGAGGTTATTAAACAGAACAGACGACCAACCAAGCTCAACACACTGCGCAAGATA 4068
QY      1171 LysGlyArgLeuAspSerSerSerGluMetAspHisSerGlu---AspGluAspTyrThrMet 1189
Db      4069 -----GAAATGGAAAGGCTCAGCGTTGAATGAAATAATTGCAAGA 4107
QY      1190 SerSerProLeuProGlyLysLysSerAspLysArgAspSerSerAspLeuValArgSer 1209
Db      4108 AGTCAGAGAGATTAATAATCTCTTAACCAAGAAAGACACCTTAACCAAGATAAAGAA 4167
QY      1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu----- 1224
Db      4168 GCCCTTGAAAGTTAAACATGACGAGCTGAAAGAACATATTAGAGAAACTTGGCTAAATC 4227
QY      1225 -----GluGluGluLysLeuGlyMet-----AspAspLeu 1234
Db      4228 CAGAGCTCTCAAGCAACCAAGAACAGCTCTTAATTAATGAAAGAAAGAACACATGAAACT 4287
QY      1235 ThrLysLeuValGluGluGln-----LysProLys----- 1244
Db      4288 ACCAAATCTGAGTACGATGAGCAATTCAAACCCAAAGATTTCAGCACTACTAGATA 4347
QY      1245 -----GlySerGlnArgSerArgLysArgGlyHis----- 1254
Db      4348 GAAATAGAAATGCTCGATTGTCTCCAAAGACTTCAGAAAGACATGATGAATGAATCT 4407
QY      1255 ThrAlaSerGluSerAspGluGln-----TrpProGluLys 1268
Db      4408 GTACCTAAGAGGAAAGATGACCTACAGAGCTGCAGAAAGTTCTTCAATCTGAAAGTGAC 4467
QY      1269 ArgLeuLysGluAspIle-----LeuGluAsnGluAsp----- 1279
Db      4468 CAGCTCAAGAAACATTAAGAAATGTAGCTTAACACCTGAAACAGCAAGAGAACT 4527
QY      1280 -----GluGlnAsnSerProProLysLysGlyLysArgGly 1291
Db      4528 AAAGTTCCTCATTTGCTGCTGAAAGAACAGAACTATTAAAGATTAAAGACTGAAT 4587
QY      1292 ArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
Db      4588 CTTTCAGAGAG-----GAACTGAAATATATCAACCATTCAAAGACAG 4629
QY      1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluGluGluGlu 1331
Db      4630 TTAGAGCAATCATGATTAATTCAGAAACAGATCCAAAGATTTATGAGAAAGGAA 4689
QY      1332 -----ArgGlnSerGlyAsnThrGluGlnLys---SerLysSerLysGlnHis 1346
Db      4690 CAACCTTAATATTAACCAATTAAGTGAAGGTTCAAGAAAGTCAAGCTGAACAAATTC 4749
QY      1347 ArgValSerArgArgAlaGlnGlnArgAla---GluSerProGluSerSerAlaIleGlu 1365
Db      4750 AAGAGCACTCGCAAGACCAAGAGATTACACACTACAAAGTATAGAAAGTAAGATGCTGAG 4809
QY      1366 SerThrGlnSerThrProGlnLysGlyArg----- 1375
Db      4810 TT-GACCAACAGACTTCAGAAAGTCAAGAAATTAACAATTATGATTAGGAAAAAGA 4868
QY      1376 -----GlyArgProSerLysThrProSerProSerGlnProLys 1388
Db      4869 GGAATGAAAGAGTACAGAGGCGCTTCAGATAGAGAGAGACCAACTGAAGAAAGAA 4925

```

```

RESULT 11
US-09-893-519A-140
; Sequence 140, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig

```

```

; APPLICANT: MOORE, Jeffrey
; APPLICANT: BURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESTIVA, Johana
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/16548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 140
; LENGTH: 7992
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLIC INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/215005.1
; DATABASE ENTRY DATE: 1993-01-10
; RELEVANT RESIDUES: (1)..(7992)
US-09-893-519A-140

Alignment Scores:
Pred. No.: 1.02e-09 Length: 7992
Score: 229.00 Matches: 285
Percent Similarity: 35.61% Conservative: 268
Best Local Similarity: 18.35% Mismatches: 546
Query Match: 3.18% Indels: 455
DB: 11 Gaps: 66

US-09-512-581B-2 (1-1391) x US-09-893-519A-140 (1-7992)
QY      17 ProGluValLysGluIleSerAspLysLysSerLysGluGluMetValAlaArgLeuLys 36
Db      997 CCTATGCTTAATGAGGTATCA-----ACTGATGAAGCTCTCGTGAAGAGTATGGA 1047
QY      37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu----- 51
Db      1048 -----AAGAAATATGATGATCTTAATAAACAATTAGAGAGGTTTCTTTAGAGACG 1098
QY      52 -----GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
Db      1099 CGGGCTCAGCAATGGAAGAAAGCCAAATTTGGCCCACTTTTGAAGAA-----AAGAT 1152
QY      67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
Db      1153 TTGCTTCAGAAAGTACAGAAATGAGAAATTAACCACTTAACACGATCTGCTG----- 1206
QY      83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
Db      1207 -----ACCTCTTCTTCC----- 1218
QY      103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
Db      1219 -----CTCAGCTTGCAA-----CAGGAATTAAGGCT 1245
QY      123 ProGlnPheAsnArgTyrPheTyrLeuLeuGlnLysHisIleAlaTyrValLysSerTyrAsn 142
Db      1246 AAAAGAAACCAAGAGCTTACTTGGCTTGGCAAAATTAACAATAAGAAAGCTCAAAC 1305
QY      143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162

```

[illegible]

QY 813 ValProAspGluGluValSerProGluThrMetValLysIleGluAlaIleLysMetMet 832
 Db 2992 -----GAGGAAATAATACAGAGAACTAAAGATGATTTTCAG----- 3027
 QY 833 ValArgThrLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
 Db 3028 ---CAAAAGATGCTGGCATGATTAATAAACAGATTTGGAAGCTTAATAATATCC---CAA 3081
 QY 853 LeuLeuThrThrIleLeuHisSerSerPclLysPleuThrGluGluGlyLysIleSerLys 872
 Db 3082 ACACAACTGACAGATGTTAAGCATATATAGATTAATGACCAACAAAGAGATATTTTCT 3141
 QY 873 -----ProAspMetSerArgLeuArgLeuAlaIleValSerAlaIleValLysLeuAla 890
 Db 3142 TTAATACAGAGAAATAATGAACTCCACAAATGTTAGAGAGTGTATATGACGAAAGAA 3201
 QY 891 Gln-----GluProCysTyThrHisGluIleIleThrLeuGluGlnTy 904
 Db 3202 CAATTGAAAGACTGACCTTAAGAAATATTGAAATGACCATTTGAATACAGAGAAATTA 3261
 QY 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyThrGlnValArgGlnValPheAlaGln 924
 Db 3262 AGACTT-----CTTGGGATGAACTTAAAGCAACAGAGATAGTTGTCACAA 3309
 QY 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyMetAlaIleCysAla 944
 Db 3310 GAAAGAAACCATGCCATTAAGAA-----GAGAGAGACCTTTCTAGACCTGTGAC 3360
 QY 945 LeuCysAla-----LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
 Db 3361 AGACTGGCAGAGATTGAAGAAATACTAAAGAAAGCAAGCCAGCACTCCAGAAATAACAG 3420
 QY 963 ValLysAsnIleAsnValArgArgLysTyThrLeuLysGlnHisAlaIleValSerLys 981
 Db 3421 CAACAACTTCTTAATGTAACAAAGAGATGAGTGCAGTGCAGAAAGATTAATTAATA 3480
 QY 982 -----LysLeuLeuSerLeuLeuProGluLysValValProTy 994
 Db 3481 GAGAAATTAAGATGATTAAGAACCAAGAAATGACATGTCGAAACATATGAAACAGAG 3540
 QY 995 ThrIleHisLeuLeuAlaHis---AspProAspTyValLysValGlnAspIle---Glu 1012
 Db 3541 AGCGTTCAGTGGCTCAGAAACTTAATGAATAATTTAGGAAGTAAATCTATTACCAA 3600
 QY 1013 GlnLeuLysAspValLysGluCysLeuThrPheValIleGluIleLeuMetAlaLysAsn 1032
 Db 3601 GAAAGAAAGTTCTTAAGGAA-----TTACAGAACTCATTTGAAACAGAG 3645
 QY 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLys 1052
 Db 3646 AGAGACCACTTAGAGATATATAGAGAAATTTAGAGTACAGCGCTTCAAAACCAAGAA 3705
 QY 1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyThrValCysAspValAla 1072
 Db 3706 GAACATAAAATTTGCTCATATTCACCTTAAGAACCAAGAAACTATGATGAAGAACTAGA 3765
 QY 1073 MetAsnIleIleMetSerLysSerThrTyThrSerLeuGluSerProLysAspProVal 1092
 Db 3766 AGAAGCGCTA-----TCTGAGAGAGACA----- 3786
 QY 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTy 1112
 Db 3787 -----GCTCAATAATTAATTAATACACAGACTTAGAAAAATCCCATACCAA-----TTA 3834
 QY 1113 ProProGluMetLysSerPhePheThrProGluLysProLysThrThrAsnValLeuGly 1132
 Db 3835 CAAGAGAGATCCCA-----GTGCTTCAT 3858
 QY 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser 1150
 Db 3859 GAGAGAACAGAGATTGCTTAATGTGAATAAAGCTAGTGAAGCTCAGAGAAACATGTAAT 3918

QY 1151 ArgMetGluThrValSerAsnAlaSerSerSerAsnProSerSerProGlyArgIle 1170
 Db 3919 GAACAGGAGTTATTATTAACAGACAGTCCAAACCAAGAGACTCAACAACTGGCAGAAATA 3978
 QY 1171 LysGlyArgLeuAspSerSerGluMetLysHisSerGlu---AsnGluAspTyThrMet 1189
 Db 3979 -----GAAATGCAAAAGCTCAGCTTGAATGAATAAATTTCAAGAA 4017
 QY 1190 SerSerProLeuProGlyLysLysSerLysAspArgAspSerAspLeuValArgSer 1209
 Db 4018 AGTCAGAGAGATTAATAATCTCTTAACCAAGGAAAGACCAACCTTAATAACGTTAAAGAA 4077
 QY 1210 GluLeuGluLysProArgGlyArgLysTyThrProValThrGlu----- 1224
 Db 4078 GCCCTTGAAGTTAAACATGACACCGCTGAAAGAAACATATTAGGAACACTTGGCTAAATC 4137
 QY 1225 -----GlnLeuGluLysLeuGlyMet-----AspAspLeu 1234
 Db 4138 CAGGAGCTCAAAAGCAACAGAACAGACGCTTAATATGAAAGAAAGAACAAATGAACT 4197
 QY 1235 ThrLysLeuValGlnGluGln-----LysProLys----- 1244
 Db 4198 ACCAAATGCTGAGTGAAGTGAAGCAATTCAAACCAAGAAAGATTCAGCACTAAAGATA 4257
 QY 1245 -----GlySerGlnArgSerArgLysArgGlyHis----- 1254
 Db 4258 GAATAGAAATGCTGCAATTTGTCCAAAAGACTTCAAGAAAGCATGATGAATGAATCT 4317
 QY 1255 ThrAlaSerGluSerAspGluGln-----TrpProGluGluLys 1268
 Db 4318 GTACCTAAGAGAAAGATGACCTCAGAGCGCTCAAGAAAGTTCTCAATCTGAAAGTGAC 4377
 QY 1269 ArgLeuLysGluAspIle-----LeuGluAsnGluAspGluGln 1281
 Db 4378 CAGCTCAAGAAATATTAAGAAATTTAGTAAACCTGGAACATGAAAGAGAACTT 4437
 QY 1282 AsnSerProProLysGlyLysArgGlyArgProPro---LysProLeuGlyLysGly 1300
 Db 4438 AAAGTTGCTCATTTGCTGCTGAAGAACAGAGAACTTAATAGAGTTAAGAGTGAAT 4497
 QY 1301 ThrProLysGluGluProThrMetLysThrSerLysLys-----GlySer 1315
 Db 4498 CTTTCAGAGAGAAAGCAATGAATATCAACCAATTCAAAGCACTGAGCAATCAATGAT 4557
 QY 1316 LysLysLysSerGlyProProAlaProGluGluGluGluGlu-----GluArg 1332
 Db 4558 AAATTACAGAACCAAGATCCAGAGATTTATGAAAGAGAACTTAATATATAACAA 4617
 QY 1333 GlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
 Db 4618 ATTAGTGAAGTTCAAGGAAACGCTGAATGATCACTGAAACAAATTCAGAGACATCGCAAGCC 4677
 QY 1353 GlnGlnArgAla---GluSerProGluSerSerAlaIleGluSerThrGlnSerThrPro 1371
 Db 4678 AAGCATGACGACATCAAAAGATATAGAAAGTATAGATGCTCGAGTT- GACCAACAGACTTCA 4736
 QY 1372 GlnLysGlyArg----- 1375
 Db 4737 AGAAAGTCAGAGAAATTAACAAATTAATATTAAGGAAAGAGAAATGAAGAGTACA 4796
 QY 1376 GlyArgProSerLysThrProSerProSerGlnProLys 1388
 Db 4797 GGAGGCCCTTCAGATAGAGAGACCACTGAAGAAATA 4835

RESULT 12
 US-10-133-013-260
 ; Sequence 260, Application US/10133013
 ; Publication No. US20030166903a1
 ; GENERAL INFORMATION:
 ; APPLICANT: Astromof, Anna
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Cocks, Benjamin G.
 ; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE


```

; FILE REFERENCE: PA-0049 US
; CURRENT APPLICATION NUMBER: US/10/133,013
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/287,067
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ. ID NOS: 271
; SOFTWARE: PERL Program
; SEQ. ID NO 260
; LENGTH: 8491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20030166903A1 441283.5
; US-10-133-013-260

Alignment Scores:
Pred. No.: 1,12e-09 Length: 8491
Score: 229.00 Matches: 285
Percent Similarity: 35.61% Conservative: 268
Best Local Similarity: 18.35% Mismatches: 546
Query Match: 3.18% Indels: 455
DB: 12 Gaps: 66

US-09-512-581b-2 (1-1391) x US-10-133-013-260 (1-8491)
QY 17 ProGlyValLysGluIleSerAspLysIleSerLysGluLumetValArgArgLys 36
DB 1087 CCTATGTATATGATGAGTATCA-----ACTGATGAAGCTCTCCGAAAGGTATAGA 1137
QY 37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu-----51
DB 1138 -----AAAGAAATATATGATCTTAAAAACAATTAAGACGAGCTTCTTAGACAGC 1188
QY 52 -----GluLysGluLeuTyrlLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
DB 1189 CGGGCTCAGCCATGCAAAAGACCAATTGGCCCACTTTTGGAGAA-----AAAGAT 1242
QY 67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
DB 1243 TTGCTTCAGAAAGTACGATGAGAAATTTGAAAACTTAACAGGATGCGGTG----- 1296
QY 83 CysLeuAlaAspIlePheArgIleTyrlAlaProGluAlaProTyrlThrSerProAspLys 102
DB 1297 -----ACCTCTTCTTCC-----1308
QY 103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuGluAspThrLysSer 122
DB 1309 -----CTCAGCTTGCMA-----CAGGAATTTAAAGGCT 1335
QY 123 ProGlnPheAsnArgTyrlPheTyrlLeuLeuGluAsnIleAlaTrpValLysSerTyrlAsn 142
DB 1336 AAAAGAAACGAGAGATTACTTGTCCTGGCCCTTGAACAAATTTAAACAAATGAGAACTCAAC 1395
QY 143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrlArgThrLeuPhe 162
DB 1396 -----TATGCAAGATCAATTT 1410
QY 163 SerValLleAsnAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
DB 1411 AATATACCAACAACAAAT-----ATAACACAAACAAACATTAAGCTTTCTTAATAATTTATTCGA 1467
QY 183 SerIle-----IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
DB 1468 GAATATGATGAATCTGCTCTTCAGAGTCTGATGATTTTTCAGTACACCTCTTGATACATTA 1527
QY 200 Leu-----ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrlAspLeuAla 218
DB 1528 AGTGAGATGAATGAGATCCAGCAACAAGCTACTAATATCAGAGGATATAGAA----- 1581
QY 219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrlIleThrIlePhePheAsn 238
DB 1581 -----1581

```

```

QY 239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValIlePheAspLeu 258
DB 1582 -----AGTGAATTGACACCATTTGCTGCTGACTGATGATAACTG 1620
QY 259 IleLeuGluLeuTyrlAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
DB 1621 GTATTA-----GACTATGACCAACTACGACAAACAAAGAAATGGA 1665
QY 279 PheLysLeuLysSerAsnAspAsnGluArgLeuGlnValValLysLeuAlaLys 298
DB 1666 TTGAATTTAAAGAAAGAAATGATTTGATGAATTTGAAGCTCTAGAAAGAAACTAAA 1725
QY 299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
DB 1726 -----AAAGATCAAGAGATGCMA-----1743
QY 319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
DB 1743 -----1743
QY 339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrlLeuLysValArg 358
DB 1744 -----CTAATTCATGAATTTGCAACCTTAAGATTTTACTTAAGCATCGAGAAATATAT 1797
QY 359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValIleAlaAla 378
DB 1798 AATCAAGATCTTGGAATGGAATC-----AGTTCA 1827
QY 379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnDheValArgGluArgThrLeu 398
DB 1828 AAAGTACAGCTGCTTACGAAAGAAAGACACAGCTTAAGAAAGCTACAGGAATACATGAC 1887
QY 399 AspLysArgTrpArgValArgLysGluAlaMetCysLeuLysGlnIleTyrlLysLys 418
DB 1888 TCTCAAAAG-----CTAGAAATATATTAATTTGACCTTGCA-----1923
QY 419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
DB 1924 TACTCATTTGGAACCATTTGAAGACCCAAACAAATGAACGACACTGTTGATGTCGAA 1983
QY 439 LysLeuLeuHisIleTyrlTyrlGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
DB 1984 ACTGTA-----GCCCTTGATGCCCAAGAGA 2007
QY 459 PheAlaGlnTyrlMetValProHisAsnLeuGluThrGluArgMetLysCysLeuTyr 478
DB 2008 GAATCAGCTTTCTTGAAGTGAATCTGTGAGTTGAAGAGAAATGAAAGAACTTGCA 2067
QY 479 TyrlLeuTyrlAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
DB 2068 ACTACATACAAAGCAATATGAA-----AATGATATTCAGTTATAT 2106
QY 499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
DB 2107 CAAGCCAAATTTGAGCGCAAAAGAAATGCAATGATCTGCGAGAAAGAA-----2157
QY 519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg-----534
DB 2158 -----TTACATATGCTTTTAAATGAGATTAACAAACCTCACCTCTTAATAGATGCC 2208
QY 535 -----AsnLeuProAspProGlyLysLysAlaGlnAspPheMet 546
DB 2209 AAAGTCCCAAAAGATTTGCTGTGTAATTTGGAATTTGGAAGAAAGATTAAGTCTTCAG 2268
QY 547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
DB 2269 AAAGAACTAATTAAGAAAGATTTGAAGAAATGAGCTTTGCGGGAAGAAAGTC---ATTTTG 2325
QY 567 ValSerProThrCysSerCysLysGlnAlaGluClyLysValArgGluIleThrLysLys 586
DB 2326 CTTTCA-----GAATTGAATCTTTAAGTGAAGTGAAGAAAGCTGAGAAAGAG 2376

```


QY 587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
::: ||| :::: ||| :::: |||
DB 2377 ATA-----CAAGACAATCTGAAAGACCTGCATATATATACATCAGAAAAAGAT 2424
QY 607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
::: ||| :::: ||| :::: |||
DB 2425 AAATGTGTTTCTGAAAGTAGTTCATAGAGAGAGTAGTTCAGAGTTTACCTTAGAGAAAT 2484
QY 626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
||| :::: ||| :::: |||
DB 2485 GGGAAACAAAAGATGACCTTACACATCAGACGTGCATTTATAAAAGCCTGATCAAGAA 2544
QY 646 Ile----- 646
DB 2545 TTCCAAAATTCCAAAACCTTCATATGACCTTTGAGCAAAAGATATAGATGCTGCTGAG 2604
QY 647 -----ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
||| :::: ||| :::: |||
DB 2605 GAGAAATGAGAGATGATCAGGAAATAGTATATCTCTCTAAA----- 2646
QY 664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp----- 680
||| :::: ||| :::: |||
DB 2647 -----GAAGCCCAAAATTTGATTCGAGCTTGGCTGTTTGAAGCCGAGCTTCTTAC 2700
QY 681 -----AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
||| :::: ||| :::: |||
DB 2701 AAGACCAAGAAGATTTCAGAGAGAAACAGCTGAGTTCCAGAAAGACTA-----AATGAG 2754
QY 696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
::: ||| :::: ||| :::: |||
DB 2755 ATGGAACAGCTGCAAGAACAAATTAGAAATAGAGATTCTCCCTGCAAACTGAGAAAG 2814
QY 716 HisLysSer----- 718
||| :::: ||| :::: |||
DB 2815 GAGAAACACAGTATTACTGACAACCTGCAGAACTTAGAAGAAATAAACTTTAACT 2874
QY 719 ---LysLysGluProArgGlnAlaLysTyraIleHisCysIleHisAlaIlePhe 737
::: ||| :::: ||| :::: |||
DB 2875 CAAGAAAAAGATGATCTAAACACTCCAGAAACCTTGCATAAT----- 2919
QY 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
::: ||| :::: ||| :::: |||
DB 2920 ---GAGAGGACCAACTCAAAAGTGTATCTCAGTACTGTTAAATGATTAATGATACT 2976
QY 758 SerAsn-----LeuGluHisLeuIleThrProLeuValThrIleGluHis 772
||| :::: ||| :::: |||
DB 2977 CAAGACAATTAACGAATGCTCTGAGTCTGGAACAACATCAAGAAACTTAATTAACA 3036
QY 773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerThrValAlaThrPhe 792
::: ||| :::: ||| :::: |||
DB 3037 CTA-----AATGCAAAATTTCTGAGGAA 3060
QY 793 IleValLysAspLeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuThr 812
::: ||| :::: ||| :::: |||
DB 3061 GTTTCGAGGAATTTGCTCATATG----- 3081
QY 813 ValProAspGluGluValIleSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
||| :::: ||| :::: |||
DB 3082 -----GAGGAAAAATACAGAGAGAACTAAAGATGAAATTTTCAG----- 3117
QY 833 ValArgThrLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
::: ||| :::: ||| :::: |||
DB 3118 ---CAAAAGATGGTTGGCATGATTAATAAACAGATTTGGAAGCTTAATAAAATATCC---CAA 3171
QY 853 LeuLeuThrThrIleLeuHisSerAspGluAspLeuThrGluGlnGluLysIleSerLys 872
||| :::: ||| :::: |||
DB 3172 ACACTAACCTGCAGATGTTTAAGAGATATAGATTAATTTGAGCAAAAGAGATATTTCT 3231
QY 873 -----ProAspMetSerArgLeuArgLeuAlaIleArgSerAlaIleValLysLeuAla 890
::: ||| :::: ||| :::: |||
DB 3232 TTAATACGAGAGAAAAATGAACTCCACAAATCTTTAGAGAGCTTTTATGACAGAAAGGAA 3291
QY 891 Gln-----GluProCysTyThrHisGluIleIleThrLeuGlnGlnTy 904

DB 3292 CAATTGAAGACTGACCTTAAGAGAAATATTGAATGACCATTTGAAACCCAGAGAAATTA 3351
||| :::: ||| :::: |||
QY 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
::: ||| :::: ||| :::: |||
DB 3352 AGACTT-----CTTGCGGATGATGAATTAATAAACCAACAGAGATGCTTCACAA 3399
QY 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyMetAlaIleCysAla 944
::: ||| :::: ||| :::: |||
DB 3400 GAAAAGACCATGCCATTAAGAA-----GAAGAGACCTTTCTAGACCTGTGAC 3450
QY 945 LeuCysAla-----LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
||| :::: ||| :::: |||
DB 3451 ACAGTGGCAGAGATTGAAGAAAACTTAAGAAAGAAAGAGCCAGCACTCCAGAAAAACAG 3510
QY 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
::: ||| :::: ||| :::: |||
DB 3511 CAACAACCTTAAATGTACAGAACAGATGAGTGAAGATGACAGAAAAGATTATAGAAATA 3570
QY 982 -----LysLeuLeuSerLeuLeuProGluTyrValValProTy 994
||| :::: ||| :::: |||
DB 3571 GAGAAATTAAAGATGATTAAGAAACAAAGAAATTGACATTGGAAACATATGGAACACAG 3630
QY 995 ThrIleHisLeuLeuAlaHis---AspProAspTyraLysValGlnAspIle---Glu 1012
::: ||| :::: ||| :::: |||
DB 3631 AGCCTTGAGTTGGCTCAGAAACTTAATGAATATTTGAGAGAAAGTGAATCTATACCAAA 3690
QY 1013 GlnLeuLysAspValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsn 1032
::: ||| :::: ||| :::: |||
DB 3691 GAAGAAAAAGTTCTTAAGGAA-----TTACAGAACTCATTTGAAACAGAG 3735
QY 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
::: ||| :::: ||| :::: |||
DB 3736 AGAGACCCCTTAGAGATATTAAGAGAAATTTGAAGCTACAGGCTTCACAAACCAAGAA 3795
QY 1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyThrValCysAspAlaAla 1072
::: ||| :::: ||| :::: |||
DB 3796 GAACATAAAATTTGCTCATATTTGACCTAAAGAAACCAACAAAGAACTATGATGAACTAAGA 3855
QY 1073 MetAsnIleIleMetSerLysSerThrThrTyThrSerLeuGluSerProLysAspProVal 1092
::: ||| :::: ||| :::: |||
DB 3856 AGAACCCGTA-----TCTGAGAAAGACA----- 3876
QY 1093 LeuProAlaArgPheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyraLeu 1112
||| :::: ||| :::: |||
DB 3877 -----GCTCAATATTAATTAATCTCAGAGCTTAGAAAAATCCCATACCAAA-----TTA 3924
QY 1113 ProProGluMetLysSerPheThrThrProGluLysProLysThrThrAsnValLeuGly 1132
||| :::: ||| :::: |||
DB 3925 CAAGAAAGAGATCCCA-----GTGCTTCAT 3948
QY 1133 AlaValAsnLysProLeuSerSerAlaLysGlnSerGlnThrLysSer-----Ser 1150
::: ||| :::: ||| :::: |||
DB 3949 GAGGAACAAGATTTCTGCTCAATTTGAAAAAAGTCATGAGACACGAGAAACAATGAT 4008
QY 1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGluArgIle 1170
::: ||| :::: ||| :::: |||
DB 4009 GAACGTGGAGTTATTTAAACAGAACAGTCCCAACCAAGAGCTCAACAAACCTGCGCAAGATA 4068
QY 1171 LysGluArgLeuAspSerSerGluMetAsnHisSerGlu---AsnGluAspTyThrMet 1189
||| :::: ||| :::: |||
DB 4069 -----GAAATGGAAGAGCTCAGCTGTGAATGAATAAAATTTTCAGAA 4107
QY 1190 SerSerProLeuProGluLysLysSerSerAspLysArgAspAspSerAspLeuValArgSer 1209
||| :::: ||| :::: |||
DB 4108 AGTCAGGAAGAGATAAATTTCTTAACCAAGAAAGAGCAACCTTAATAACGATTAAGAA 4167
QY 1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu----- 1224
||| :::: ||| :::: |||
DB 4168 GCCCTTGAAGTTAAACATGACACAGCTGAAGAAAGACATATTAAGAAACTTTGGCTAAATC 4227
QY 1225 -----GlnGluGluLysLeuGlyMet-----AspAspLeu 1234
||| :::: ||| :::: |||

OY		283	serasn-----	-aspangluglurgleuqlnvalylals	294
Dd		1931	GCCACGGGTATCATCTGAGAGAGACTGCTGGCTTGACCACAAACACATGCTCACCGAG	1990	
OY		295	LeuLeuAlaLysMetPheGlYAlaLysAspSerGIuleuAlaSerGlnAsnIysProleu	314	
Dd		1991	TTTTTCTCTCCTCACTGATC-----GCACACTGGTTCTG	2022	
OY		315	TrrGlnIysTYrLeuGlYArgPheAsnAspIleHisValProIleArgLeuGluCysVal	334	
Dd		2021	TGGTTCCTGCCACACTGGGGCTT-----CTGCCCGAAGGCTTGATGATGCTCT	2065	
OY		335	LysPheAlaSerHisLysCysLeuMetAsnHisProAspIleAlaLysAspLeuThcGIunTYr	354	
Dd		2066	CTGCACCTGTCAAGACTACGCTACACACTACCTCTGTAATTCTACAGACTACTATACAAAGTGGT	2122	
OY		355	LeuLYsValArgSerHisAspProGIunAlaAlaIleaRHisAspValIleValSerIle	374	
Dd		2126	CGCGCGAAGTGTGAAGCGCTCACCCGAGAGGAAGTAGAACCCAGAC-----	2170	
OY		375	ValThrAlaAlaLysLYsAspRIleuLeuValAsnAspHisLeuLeuAsnPhenValArg	394	
Dd		2171	-----TGCTTGATGATTGATGGATTATGCTTCACACTAGCA	2206	
OY		395	GluArqThrLeuAspLYsArgTrPARgValArgLYsGluAlaMetMeGIY-----	411	
Dd		2207	GATGAAA--GTTGAGAAGTGTGTGGAG--CGGAGCAAGAGTGGCACAGAAATACCT	2262	
OY		412	-----LeuAlaGlnIleTYr-----LYS	417	
Dd		2261	CCTGCTAGACATCCACGCACTTACTTGATATATATTTATTAGACAACATTACCGAAGCAAGC	2320	
OY		418	LYSTYrAla-----LeuGlnSerAlaAlaGLYLYs---LEU	427	
Dd		2321	AAACATGATTTACCATTTATTTCGTGCTGTGATATATATATTCCTTCCAAATRAAACG	2380	
OY		428	-----AspAlaAlaLysGlnIleAlaTrPIleLysAspLYsLeu	440	
Dd		2381	GATACCCCCCATTTGAATCTTCCCACAGCTCTTGGTATATTTCTGGGGCCAAGTTAAAGCTA	2440	
OY		441	Leu-----HisIleTYrGlnAsnSerIleAspAspArgLeu	453	
Dd		2441	GTTCAAGATTTTGGCTAGCTAGATCATATAGCTATATGAGAAATGGTTTAC-----CTT	2494	
OY		454	LeuValAlaIuaRgIlePheAlaGln-----TYrMetValProHis	466	
Dd		2495	CTGTTTGACACCCATGTACTGCAAAACCTGCATCGTGGCAACATTCAAAGATTAATTGAAGCT	2554	
OY		467	AsnLeuGIunThrHngIuaRgMeLysCysLeuTYrTYrLeuTYr-----	481	
Dd		2555	TTTATGAGTCAAGGAGAGCACAAACAGGCTCTCCGGTATCTTCAGACANTGAAGCCAACA	2614	
OY		482	-----AlaThrLeuAspLeuAsnAlaVal-----LYsAlaLeu	492	
Dd		2615	GTCGCCAGTACCAATGAAGATTATCTTCACCTCACTGTTTCTACACTTTTAAATAGATGATCG	2674	
OY		493	AsnGluMetTrpLYsCys-----GlnAsnLeuLeuArgHisGlnValLYsAspLeuLeu	510	
Dd		2675	GTTTAGGCTGGAACCTTACTGCGACAGATAATTCAAACAGATTAATATAGAGAAATTTATTA	2734	
OY		511	-----AspLeuIleLYsGlnPro	516	
Dd		2735	AAGCAGCGTTATGAAAGTTTGTCACAGAGATGGGCTTAATGAGAGATTTACTAAGCTGCCA	2794	
OY		517	LysThrAspAlaSerValLYsAlaIlePheSerLYsValMetValIleThrArgAsnLeu	536	
Dd		2795	TTTTCACAAACACTGAG-----	2809	
OY		537	ProAspProGluLYsAlaGlnAspPheMetLYsLysPheThrGlnValLeuGluAspAsp	556	
Dd		2810	-----CAGGATTCGCTTAATGAAATTT-----TTACAGTCCAGT	2842	
OY		557	GluLYsIleArgLYsGlnLeuGluValLeuValSerProThrCysSerCysLYsGlnAla	576	

Db	2843	ACCAAGTGTGGAAATCATGAAATTCCTCTCTACT--TCACCAATTTTACAGCTGC--CAATTA	28998
Qy	577	GIUGIYCSValArg-----:::	581
Db	2899	TATTTCTCCCTTGGAAACTAAACCAAGATTCTGAAGATTAATCTCATGATGATTCGGACCC	2958
Qy	562	GIUlethrLysLysLeuGIYAsnProLysGlnProThrAsnProPheLeuGlu-MetII	601
Db	2959	TCGATTTGGGGAAAGATCGGTGACACGCAAAATTCATATATTAGACAGATGTGGGAAATTCCT	3018
Qy	601	e-----LysPheLeuLeuGIArgIleAlaProValHisIleAspThrGI	616
Db	3019	ACCTAGATCCAGAAAGATTAGCTGTTGACGACGACTAAAGCTTACACCTGTGTGACATC	3078
Qy	616	userIle-----SerAlaLeuIleLysGlnValAs	626
Db	3079	CTCAGATTTTTCATGAAAGTTTCTAGACCCCAACCGTTATTCGGCATTTTCCCAAG--	3130
Qy	626	nlYSerIleAspArgIThrAlaAspArgIAspGluIAlaPro-----	641
Db	3131	AAAGCTATTAAGTGGAAAGCTTTAACCSSGATCTACGTTCAATCAGCAATGTTTATCTAA	3189
Qy	642	-----ThrAspGlnAlaIleArgAlaGluLeuLeuLeuLys-----	654
Db	3190	AATTCGACAGGTTGGCGCAAGCTCATGACAGCTTAAGAAATCGCGTCATCTTTTAAACAGTCC	3249
Qy	655	-----ValLeuSerPheThrHisProIleSerPheHisSe	666
Db	3250	TAAACAGACAGCCATCTCTCTAGTACACTCTTCCACACCCA-----	3295
Qy	666	ralAGIuThrPheGluSerLeuAlaGluLeuLysMetAspArgIuLysValAlaG1	686
Db	3296	-----GAGCTCTCTCGA	3306
Qy	686	uAlaAlaLeuGlnIle---PheLysAsnThrGlySerLysIleGluGluAspPheProH1	705
Db	3307	GCGCTGTCTGGAACTCCCAATTTCAAAAGCTCCACAGAAAT-----	3349
Qy	705	slleArgSerAlaLeuLeuProValLeuHisLysSerLysGluProProAlaG1	725
Db	3350	-----CTTAGATTAAGTGGATTTGGTTGGTGCAT-----CCTGTAACCCA	3387
Qy	725	nalAluLysTyrAlaIleHisCysIleHis-----AlaIlePhe	737
Db	3388	GCGTTCACAGTGTGGAGTTTAAACAAAGTCCCAAGATCTCTTGTTGCTGCT	3447
Qy	737	eSerSerLysLysIuThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspRg	757
Db	3448	GTCACAGTCTGTTACCAATTAAGTCAAGCTTTAAAGGCGACATCAGATTAAGTCCAGGCC	3507
Qy	757	oSerAsnLeuGlnHisLeuIleThrProLeuValThr-----IleGlnHisIleAlaLe	775
Db	3508	TTTCAGAGTTCGTTTACTTACTGAGACTCTCTCATGATTAAAGAACTTAATCTTGGCTGCT	3567
Qy	775	uLeuAlaArgAspGlnPheAlaAlaArgProGlySerTyrValAlaIleThrPheIleValL	795
Db	3568	GTCAGCCSACG-----TCTTCGGATTTTCCSAGATTCTACTCTCC	3606
Qy	795	sAspLeuLeuMetAsnAspArgLeuProGluLysThrThrLysLys-----	811
Db	3607	ATGCATCTCTAGTCT-----GCTTTTGCAAACAAACACTTTTAGCATCTCTCTC	3654
Qy	812	-TyrValProAspGlnGluValSerProGluThrMetValLysIleGlnAlaIleLysMe	831
Db	3655	TTTTCTGACCTGGAAAGATCTCTACCTCCGCTTTCAGAGTTTAAAGAAACAAAGATTTTCAT	3714
Qy	831	tMetValArgThrLeuLeuGlnMetLysAsnAsnHisSerLysSerGluThr-----	848
Db	3715	CATGGAGAA-----GCGATGAATATACACATGATAGATATAGACATACAGATGACCG	3765
Qy	849	SetThrLeuArgLeuLeuThrThrIleLeuHisSerAspArgL-----	862

Db 3766 AATATACAAAAGCGTTTGTAGACACATCTTTCCATTAATGTGACCTTCAGACAGAACTGA 3825
 QY 863 -----Aspleu----- 864
 Db 3826 GTGGATGAAGACCAGTGAATACATATTTCTCTGGATGTCCCTGCAGAAAGGCC 3885
 QY 865 -----ThrgluInglYlSerLysSerLysProAspke 875
 Db 3886 TCAGAAAGTGTGGCAGAGTCACTGGCTACCCATTCAGAGAGGCTGAGAAAGCATGATGT 3945
 QY 875 tSerArgLeuArgLeuAlaLysSerAlaLeuAlaLysLeuAlaGluProCysIy 895
 Db 3946 GAGCAAGAAGACAGACAGCTTCCACCGTCAGACAGACCTCCTTAGAG-----TA 3999
 QY 895 rHsgLuiLeIethrLeuGluInLyrGluInLysAlaLeuAlaIle----- 911
 Db 4000 TCATGCGCACCATCCAGAAAGACTTGGAAAG-TCTGTGTTTTGTGTCTCCCAAGCCAG 4058
 QY 912 -----AsnAspGluCysTyrGlnIya 918
 Db 4059 CATCTTCCTCCACTGAATACTAATAATTCACCTCAACACAGAGAGGG-ATATATAT 4117
 QY 918 lArgGlnAlaPheAlaGlnLys-----LeuHis----- 927
 Db 4118 AAGATGCGTTTAAAGCAGAAAGGTACTCCTTCACCGTGAAACAAATAGCGACGGGA 4177
 QY 927 ----- 927
 Db 4178 GACGTCGACGTGAAGCATTTTCAGAACTGAGTCGCTTAGACCCCTGTGAAAGAGCTGAA 4237
 QY 928 -----LysGluLysSerArgLeuArgLeuProLeuG 938
 Db 4238 GCTTCTTTGGTGTGTGTCAGTCTGTGAAGGGGAAACCTCCACTCAAACTCCAGAGAG 4297
 QY 938 u-----TyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg----- 954
 Db 4298 TCAGTTCTGATGGAATCGTGCCTATTGAGAGCCGAACCTCACTTACAGCAGACAC 4357
 QY 954 ----- 954
 Db 4358 AAAGAGTCTGTGCGCCAAACAGCGTTCAGATGTTGAAGCTCTGGGTCCACAGCTCCAA 4417
 QY 955 ----ArgAlaHisAlaArgGlnCysLeuValLysAsn-----11 966
 Db 4418 TGCCCGGTTAACTCTGAAGCAGCGCTCGGCGAAAACTAAATTAACCTTAAGAAAGAT 4477
 QY 966 eAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLe 986
 Db 4478 GAAATGAAGCTCATGTACCAAGGAGAACGTTGGTTTACACAGAAAGAGCCCTCGAAT 4537
 QY 986 uLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa 1006
 Db 4538 TCTGGCTGCTCTCTGATACT-CACGAGATTCATCTAATTTGGA-----TGTGAATCT 4590
 QY 1006 lLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuG 1026
 Db 4591 TGAAGTTCAAAATTCGAGAGAG-----GAGGCCAAGAACTTTCTATTGATGAGTTGTA 4644
 QY 1026 uIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysMetValGluAs 1046
 Db 4645 TCCCTTAGGGGCGAGAAACTTGATATATCTCACTACTATTGAGCAGAGAGTTTGTGA 4704
 QY 1046 nIleLysGlnThrLysAspAlaGlnLyrProAspAspAlaLysMetAsnGluLysLeu 1066
 Db 4705 CTTCGCTGATGACAAAGACTCTGCTGAATGTATGTGCTGAAGTGAAGCGGAGACTTTT 4764
 QY 1066 rThrValCysAspValAlaMetAsnIleLeuMetSerLysSerThrThrTyrSerLeuG 1086
 Db 4765 T-----GTGGCCAGAGCAACTTAACTTACCTGATTTTGA 4797
 QY 1086 uSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnHese 1106
 Db 4798 AGGTGAAGAAGA-----GAAGCTGAGGCAAGCGACCTC 4830

QY 1106 rAsnThrLysAsnTyrLeuProGluLysSerPhePheThrProGluLysPro-- 1125
 Db 4831 TGCAGACCTAATATGTTATCCGAATACGCAAGAA-----AAACCTGT 4875
 QY 1126 -----LysThrThrAsnValLeuGluAlaValAsnLys 1136
 Db 4876 GTGCTACAGGAACCCCATATATCAGAGCGCGTTCAGATTTGCCATGCTGCTGACT-- 4933
 QY 1136 sProLeuSerSerAlaGluLysGlnSerGlnThrLysSerSerArgMetGlnValSe 1156
 Db 4934 -----GCTACCAAGATCCCAACAGGTGAGACTTTA-- 4966
 QY 1156 rAsnAlaSerSerSerAsnProSerSerProGluArgIleLysGluArgLeuAspSe 1176
 Db 4967 -----CCGTAATGTGCTGGAACCGGTTAAAGTGGCAATTGCCA 5004
 QY 1176 rSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlu 1196
 Db 5005 AAATCTGTGGATGTAATTAAGACACACAGAGTAAGAGCAACTCCCGTGACAGCAG 5064
 QY 1196 sLysSerAspLysArgAspAspSerSerLeuValArgSerGluLeuGluLysProArg 1216
 Db 5065 TGAGCGTGTGATGACAGCAGCAGTGTATGTCTAAAGCTGCACATTCGTCCAGGCT 5124
 QY 1216 yArgLysLysThrProValThrGluGlnGluLysLeu-----G 1230
 Db 5125 GACAACTCTACAGCCGAAGACTGTTAAGAACCCGTCGAGAGACTGTAAATACAGCA 5184
 QY 1230 yMetAspAspLeuThr----- 1236
 Db 5185 GAGTGATGACATGCTTTCTTGTAGACTCTCAAGAAAGCAGCATGCCCTAAGCTGAA 5244
 QY 1236 sLeuValGlnGluGlnLysProLysGlnSerGlnArgSerArgLysArgGluHisThrAl 1256
 Db 5245 TGTCACTAGACAGAACAGCCTTCAGCACTGTGCCACTCTTAAGAAAGACTAGAAAT 5304
 QY 1256 aSerGlu-----SerAspGluGlnIleTyrProGlu-- 1266
 Db 5305 TAAAGAACTCCGAGACTCTTGAAGAGACCTGTGTGACCTAAAGTAGACACTGAGAA 5364
 QY 1267 -----GluLysArgLeuLysGluAspIle----- 1274
 Db 5365 CCAACTGACAGCTCAAAATCTCCCGCTCTGAGGAGCGGAGCCTCAGGGTACGCCA 5424
 QY 1275 -----LeuGluAsnGluAspGlu-----GlnAsnSerProPr 1285
 Db 5425 AGGCACACTGCCAAGTTCTGTGTGCTGTGGAGCCGGAGCCTCAGGGTACGCCGG 5484
 QY 1285 oLysLysGluLysArgGluArgProProLysProLeuGluGluGluTyrPro---LysG 1304
 Db 5485 AAGACTGAGGCTGTGAGAACGAGCAGCCGAGCAGAGCTGAGAAACATCTCTGTGAAC 5544
 QY 1304 uGluProThrMetLysThrSerLysLysGlnSerLysLys----- 1318
 Db 5545 AAAAGTCAGGCTTTCATCTGTAGAAAGGGAACCCCTAAGAGACTTAAGAGCTGTAGA 5604
 QY 1319 -----serGluProProAla----- 1323
 Db 5605 AAATGGCAAAAGTATGAATTTCTAGATGATCTCAAAAGGAGTGAAGCAGCAATCATGA 5664
 QY 1324 -----ProGluGluGluGluGluArg 1332
 Db 5665 CGGAGCTGCACAGACTGAGAAATGCCAATTTAGAAATATCTACAAATATGAGATATA 5724
 QY 1332 gGlnSerGluAsnThrGluGlnLysSerLysGlnHisArgValSerArgArgAl 1352
 Db 5725 ACAAGATGAACACAGAGCAGCAACCGCTCTAAACCAAGAGAGGTGAGAGAGAGA 5784
 QY 1352 aGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrGln----- 1368
 Db 5785 ACTTACTGTGTCAAGTGTGACAGAAAGCCAAAGCTTGACTCATCTCCAGTGTGCTCTTCA 5844

```

Oy      1369 -----SerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db      5845 GACAGACTGCATGTACTGCCACCCCTAGAGAA---CCTGGTAGACCCAGAGAG 5896

RESULT 14
US-09-882-227-623
/ Sequence 623, Application US/09882227
/ Publication No. US20030158396A1
/ GENERAL INFORMATION:
/ APPLICANT: Kleantous, Harold
/ APPLICANT: Al-Garawi, Amal
/ APPLICANT: Miller, Charles
/ APPLICANT: Tomb, Jean-Francois
/ APPLICANT: Oomen, Raymond P.
/ TITLE OF INVENTION: Identification of Polynucleotides
/ TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
/ FILE OF INVENTION: Genome
/ FILE REFERENCE: 06132/047002
/ CURRENT APPLICATION NUMBER: US/09/882,227
/ CURRENT FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 08/902,615
/ PRIOR FILING DATE: 1997-07-29
/ NUMBER OF SEQ ID NOS: 638
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 623
/ LENGTH: 5334
/ TYPE: DNA
/ ORGANISM: Helicobacter pylori
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (22) .. (5250)
US-09-882-227-623

Alignment Scores:
Pred. No.:      1.25e-09      Length:      5334
Score:          225.00      Matches:      285
Percent Similarity: 32.49%      Conservative: 217
Best Local Similarity: 18.45%      Mismatches: 523
Query Match:    3.13%      Indels:      520
DB:            12      Gaps:      71

US-09-512-581b-2 (1-1391) x US-09-882-227-623 (1-5334)
Oy      4 SerLysThrArgThrAsn---AspGlyLysIleThrTyProProGlyValLysGluIle 22
Db      955 ACAGAACTCAACCAATTTTGCATGACAGACAAGTA-----GACAAATA 999
Oy      23 SerAspLysIleSerLysGluGluMetValArgArgLysLys----- 36
Db      1000 ACTGACGACTCTACGATCAAGAGATGTATCAAGAGCAAAAAGAAATATATTGTG 1059
Oy      37 -----MetValValLysThr 41
Db      1060 GGCATTGTATCGCTGTTCTTATCGTATATTATTTTCTAGAACATTTTTCACCTAC 1119
Oy      42 PheMetAspMetAspGlnAspSer-----GluGluGluLysGluLeuTyLeuAsn 58
Db      1120 TTCATGCTTTGGAGATGAAGATGCTCTGCTTTTACCAAGACAGGAATCTTTATGTCAT 1179
Oy      59 LeuAlaLeuHisLeuAlaSerAspPhe-----PheLeuLysHisProGlyLysAspVal 76
Db      1180 GATGAATCCAAATATAGGCAAGATATACGATGCTGAAAGAAAGGAAATGAAGAAAGGC 1239
Oy      77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyAlaProGluAlaPro 96
Db      1240 AATATGATCGAT----- 1251
Oy      97 TyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116
Db      1252 -----AACAATCTTTTCTTC----- 1266
Oy      117 LeuGlnAspThrLysSerProGlnPheAsnArgTyrPheTyLeuLeuGluAsnIleAla 136

```

```

Db      1267 -----AATGACGATCCCAATAGAACCTTATACAACTATTTC-----AATATTGCA 1311
Oy      137 TrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGln 156
Db      1312 -----GAAATTGAGAGACAAAACCCGTTG----- 1335
Oy      157 LeuTyArgThrLeuPheSerValIleAsnAsnGlnHisAsnGlnLysValHisMetHis 176
Db      1336 -----AAGACCTTTTATGATATGTATTACTTAATGGTGGCACTATGAAAGATGTTTGAA 1389
Oy      177 MetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeu 196
Db      1390 CTATATC----- 1395
Oy      197 AspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyAsp 216
Db      1396 -----AAAGCAAAAACCTTCAAGATCAG----- 1419
Oy      217 LeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyIleThrThrPhe 236
Db      1420 -----ATGAAAAGACTCTAGAGCT----- 1440
Oy      237 PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPhe 256
Db      1441 TATAACGACTGCATCAAAAATGCCAAACT----- 1470
Oy      257 AspLeuIleLeuGluLeuTyAsnIleAspSerHisLeuLeuSerValLeuProGln 276
Db      1470 ----- 1470
Oy      277 LeuGluPheLysLeuLysSerAsnAspAsnGluArgLeuGlnValValLysLeuLeu 296
Db      1471 -----GAAGAGAAAGATCAACTGTTTACATTTATTC 1503
Oy      297 AlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyProGln 316
Db      1504 -----AAAGATGAAAACCTTA-----AAAAAAGCTTA----- 1530
Oy      317 CysTyLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPhe 336
Db      1531 -----CTGACCCAAACAAAAGCTTCAAGTGGCGCTAGATGTTGTTGAAAAC 1575
Oy      337 Ala-----LeuMetAsnHisPro----- 345
Db      1576 GCTAAACCGATGAAGAAAGCAAGACGAGTCCATAACATCAATAAGACCGTGAAGTTAGA 1635
Oy      346 -----AspLeuAlaLysAspLeuThrGluTyLeuLysValArg 358
Db      1636 GAGAAATTCGCTAAGGAATTAAGAGCTTCAAAAGAGCTTCAAGAGTAT----- 1683
Oy      359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
Db      1684 -----AAGATTGTATCAAAAACGCCAAACACAGAACT 1716
Oy      379 LysLysAspIleLeuLeu-----ValAsnAspHisLeuAsnPheValArgGluArg 396
Db      1717 GAGAAAACCAATGCTTGAAGAGCTTGCTTAAGAAGCTATATAGAGATTTGAACAGCAA 1776
Oy      397 ThrLeuAsp-----LysArgThrArgValArgLysGluAlaMetMetGly 411
Db      1777 GCGTAGATGTTGTTGAAGAACGCTTAACCCGATGAAGAACGAAAGAGAGTGTGAAAAAT 1836
Oy      412 LeuAlaGlnIleTyLysLysTyAlaLeuGln----- 422
Db      1837 ATTCGCCCAAGACTGCAAAAAGAACTATTACTGTATAGACGTCGAAGCTTACAGAGAT 1896
Oy      423 -----SerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
Db      1897 TGCATATCAAAAGCTAGAAATGAAGAAAGCAAGAAACAGATC-----GAGCAATTTG 1947
Oy      441 LeuHisIleTyTyGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
Db      1948 CTCACGCTCGAAGGAGAGAAAAGTTAGAACACACAGTTCTA----- 1989

```

OY	461	GlnTyrMetValProHisAsnLeuGluThr--ThrgluAargMetLysCysLeuTyr	479
Db	1990	-----GATGTTTGGAAAAACGCTAAAGCCGATGACAGAACAAAAAGTGTGAAAGAT	2043
OY	480	LeuTyrAlaThrLeu-----AspLeuAsnAlaValLysAlaLeuAsn-----	493
Db	2044	CTCCCTAAAGACTTATACAAAGCGATATTTATCCCAAGAGACCGTAAAGCTTATAAAGAC	2103
OY	494	-----GluMetTrpLysCysGlnAsnLeu-----	502
Db	2104	TGCGATTCAGAACCCAAACCGAGCTGAGAAAAAAGATCGGAGAAATTACTACCCCT	2163
OY	503	-----ArgHisGlnValLysAspLeuAspLeu	512
Db	2164	GAAGCGAAAAAATCTTTAGAGAAAGAACCCAAAGACGACCGCTTATTAAGCTTC	2223
OY	513	IleLysGlnProLysThrAspAlaSerAlaLysAlaIlePheSerLysValMetValIle	532
Db	2224	GTATCTCAGCGCAAAACCGAAGCTGAGAAAAAAGAAATGC---GAGAAATTTCTCAC--	2277
OY	533	ThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysPheThrGlnVal	552
Db	2278	-----CCTGAAGCGAAAAAAGATTAGAGAAAGCTTAAAAAAGCGTTAAAGCT	2325
OY	553	LeuGluAsp-----AspGluLysIleArgLysGlnLeuGluValLeu	566
Db	2326	TACTTGGATTCCGATCAAGACAGTACGAAATGAAAAAGAAAAAGAAATGCGCAAAATTC	2385
OY	567	ValSerProThrCysSerCysLysGlnIleGluGlyCysValArgGluIleThrLysLys	586
Db	2386	CTCACCCCTGAAGCGAAA---AAACTTTAAGACACACAGACACTAGATTGTTTGAAAAAC	2442
OY	567	LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPheLeuLeuGlu	606
Db	2443	GCTAAACCCGATTAAGAGCAAAAAAGCTTTGAAGACATCTCCCTAAAGACTTCAGAAA	2502
OY	607	ArgIleAlaProValHisIleAspThrCysIleSerAlaLeuIleLysGlnValAsn	626
Db	2503	AAGGTT-----TTAGCTAAACAGCGCTTAAGAGCTTACTTGATGGCTGATCT	2550
OY	627	LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIle	646
Db	2551	CAAGCGCAAACTGAAGCTGAGAAAAAAGAAATCGAGAAATTACTACCCCTGAAGCG---	2607
OY	647	ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer	666
Db	2608	AGAAAACCTTTAGAGAGAACTAAA-----	2631
OY	667	AlaGluThrPheGluSerIleLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu	686
Db	2632	-----AAAAACGTTAAAGCGCTTATTTGGATTGC-----GTATCTCAA	2667
OY	687	AlaAlaLeuGlnIlePheLysAsnThrCysLysIleGluGluAspPheProHisIle	706
Db	2668	GCCAAACACTGAAGCTGAGAAAAAAGAAATCGAGAAATTACTACACC-----CGTGAAGCG	2721
OY	707	ArgSerAlaLeuLeuProValLeuHisLysSerLysLysGlyProIleArgGlnAla	726
Db	2722	AGAAAACCTTAA-----GAAGAAAGCT	2742
OY	727	LysTyrAlaIle-----HisCysIleHisAlaIlePheSerSerLysGluThr	742
Db	2743	AAAGAGACCGTTAAAGCTTATTAAGACTGCGCTATCAAAAGCTAGAGAAAGAAAGAGAAA	2802
OY	743	GlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHis	762
Db	2803	AAA-----GAATGGAGAGAA	2817
OY	763	LeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAla	782
Db	2818	TTACTACGAGCT-----GAAGCGAAAAAACTTTTAGACACAA---2856	

QY	783	AlArProTPrLySerThrValAlaThrPheIleValLysAspLeuIleMetAsnAspArg	802
Db	2857	-----GTGGTAGATGCTGTTGAAAAAGCGTAA	2883
QY	803	LeuProGluLysThrThrLysLeuThrValProAsp-----GluGluValSerPro	820
Db	2884	ACCGAAGCTGATTA-----AAAAGGTGTCAAAAGATCTCCCTTAAGACTTGCAGAA	2937
QY	821	GluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeuGluMetLys	840
Db	2938	AAGGTTTTCAGCTAA-----GAGAGCGTTAAAGCGTTATTGTGACGTGATCAAGAGCTAAG	2994
QY	841	AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThrIleLeuHisSer	860
Db	2995	AATGAAAAAGAGAAAA-----	3012
QY	861	AspGluAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuAspLeu	880
Db	3013	-----GAATGGCAGAAATTTGCTCAACCCCTGAAAGCGAAAAAATTTTGAA	3057
QY	881	AlaAlaGlySerAlaIleValLysLeuAlaGlnLysProCysTyrGlnHisGluIleIleThr	900
Db	3058	GAAGCCCAAGAGAGCTT-----AAACCTTATTAAGACTGCGCTCTCAAGCTAAGAAAT	3111
QY	901	LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln	920
Db	3112	GAAGAGAGAAAGAGAGCTTGGCAGAACTACTCAGCGCTGAA-----CGCAGAGAAA	3162
QY	921	ValPheAlaGlnLysLeuHisLysGlyLeu-----SerArgLeu	933
Db	3163	CTCTTAGAGCAAGAAATTAAAGAAAGCGTTATTGAGCTGCGATTAAGAGCT	3222
QY	934	ArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu	953
Db	3223	ACGAAATGAAAAAGCAAAAAAGAAATGCCAGAAATTAATCTCAGCGCTGAAAGCGAGAAATTT	3282
QY	954	ArgArgAlaHisAlaArgGlnCysLeuValLys-----AsnIleAsnValArgArgGlu	971
Db	3283	TTAGCGAAGCAAGTGCCTAAATTGTTTGAAAAAAGCTGAAATGAAAGAAAGAAAGCA	3342
QY	972	TyrLeuLysGlnHisAlaIleValSerGluLysLeuLeuSerLeuLeuProGluTyrVal	991
Db	3343	TGCTCTTAAAAAT-----CGACGAACTTAAGAAAGCGTTAAAG	3369
QY	992	ValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIle	1011
Db	3370	CAGGAA-----AATATTTTTCAGCTAA-----	3390
QY	1012	GluGlnLeuLysAspValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLys	1031
Db	3391	GAGAGCTCTTAAAGCTTATTAAGAGCTGCTC-----TCTCAAGCTAGA	3432
QY	1032	AsnGluAsnAsnSerHisAlaPhe-----IleArgLysMetVal	1044
Db	3433	AATGAAAGAAAGAGAGAGCTTGGCAGAGAACTACTCAGCGCTGAAAGCGAGAAAACTCTTA	3492
QY	1045	GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys	1064
Db	3493	GAG-----CAAGAGCTTAAGAAAGCGTTAAAG	3519
QY	1065	LeuTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer	1084
Db	3520	GCTTATTTTGGAGCTGGTATCAAGAGCTAAGATGAAAAAGCAAAAAAGAAATGCGAGAAA	3579
QY	1085	LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn	1104
Db	3580	TTACTCACAGCGCTGAAGAGAGAAATTTTTCAGAAAGAAATCTCAACAAAGAAAGATTAAGCG	3639
QY	1105	PheSerAsn---ThrLysAsnTyrLeuPro-----	1113
Db	3640	ATCAAAAGATGCTTGAAGAAAGCGCGATCTTAACAGACAGAGCGCTATCATGAAGTGTGG	3699
QY	1114	-----ProGluMetLysSerPhePheThrProGluLysProLysThrThr	1128

```

Db      3700 GATGTTTGAGCATGAGAAAGCTCAATACCTCAGAAAGCTAGAGAAAAGCTGTT 3759
      1129 AsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysLeuIntrLys 1148
      3760 GCGGATTTGTTGGCTATAGCTAAACCGATGAAGAAAAGAAAGCCAAAACCTTAT 3819
      1149 SerSerArgMetGlyThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly 1168
      3820 ACCGATTTGATCCAGAAATCCAAATTAAGACACCAAAACCAAAATCAATGAGT 3879
      1169 ArgIleLysGlyArgLeuAspSerSerGlu---MetAspHisSerGluAsn-----Glu 1185
      3880 AAAACAGAAAGGTTGCATCAGCAACGAGTGGCTGTGATTAACCTAGATGACCTACTGAT 3939
      1186 AspTyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg----- 1201
      3940 CAAGAGCCATAGACCAATGTTTAGAGGCTTGAGCGATAGTGAAGGCGCTAATTCAT 3999
      1202 -----AspAspSerAspLeuValArgSerGluLeuGluLysProArg 1215
      4000 GGAATTAAACGACAGCTGATGAGTGGATCTGATTTATAGCATTAAGAAACCGTAA 4059
      1216 -----GlyArgLysLysThrProValThrGluGluGluLysLeuGly 1230
      4060 ACCTTTGATACATGGCGCTAAGGATGATTCATTTGATCAATGATTTCAAAATGGC 4119
      1231 MetAsp-----AspLeuThrLysLeuValGluGluGln 1241
      4120 GCGCATATTGCCACTATTAAACCCCACTAATGTTGAGCGGCAAAATACCTAGCCATAT 4179
      1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp--- 1260
      4180 CCTATT-----TATGCTTCCATAAGCCTGATATT 4209
      1261 GluGlnGlnTTPProGluLysArgLeuLysGlyAspIleLeuGluAsnGluAspGlu 1280
      4210 GCCAGCAATACGAAACAGAAACCAATTAAGATGAAGAAATTAAGA----- 4257
      1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGly 1300
      4258 -----GCTAAATTAAGCTAAGGCTTATGGTGCAT 4287
      1301 ThrProLysGluGlnProThrMetLysThrSerLysGlySerLysLysSerGly 1320
      4288 -----AAAAAGATGACGATTAAGAAAAAGTAAAAATCCACACACAGCTAAAGCA 4341
      1321 ProProAlaProGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
      4342 GAAACCAATTAAGATGACAAAGATGCCGAGAAACGCAAGAAATATCAGTGAATCGCT 4401
      1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlu-----Ser 1358
      4402 CTTTAGAACAAAAAGAAAGAGTGGGAATTTAGTAGAAATGTAATCCCATGAT 4461
      1359 ProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLys-----Gly 1374
      4462 GACAAAGAAAGAAAGCAAAACAGATGAACAGCCCTGTCAAAAGCCCTTTATAGCC 4521
      1375 ArgGlyArgProSer 1379
      4522 AAGAGTATGCCACA 4536

```

RESULT 15

```

US-10-171-581-49
; Sequence 49, Application US/10171581
; Publication NO. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes In Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999

```

```

; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 49
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: Homo sapiens
; DATABASE INFORMATION:
; DATABASE ACCESSION NUMBER: D21262
; DATABASE ENTRY DATE: 2001-06-18
; US-10-171-581-49

Alignment Scores:
Pred. No.: 4,26e-08 Length: 3727
Score: 205.50 Matches: 141
Percent Similarity: 33.64% Conserved: 77
Best Local Similarity: 21.76% Mismatches: 256
Query Match: 2,86% Indels: 175
DB: 14 Gaps: 26

US-09-512-581b-2 (1-1391) x US-10-171-581-49 (1-3727)
      858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
      115 CTAACAGC-----AGATGCCAATGCTCTTCCCTCTTGG 150
      878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnLysProCysTyrHisGlu 897
      151 ACATCATAGCTTCTGCTCAAGTCTGCCAAGG---TCCACAGCGCAAGTTACAGAGCA 207
      898 IleIleThrLeuGlnGlnLysGlnLysCysAlaLeuAlaIleAsnAspLysLysTyrGln 917
      208 -----ATGACCAGTGG---CTAAGAAAGCTAAGAAAGGCTCATCCAGTGACA 255
      918 ValArgGlnValAlaPhe---AlaGlnLysLeuHisLysGlyLeuSerArgLeuArg----- 934
      256 GTAGGACAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 315
      935 ---LeuProLeuGlnLysMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu 953
      316 TACCTGCCAAGCAGAG-----TCGGTCTCCCTCTGGGAAGGCTGCAGCCAAAGCATTCAG 369
      954 ArgArgAlaHisAlaArgGlnCysLeuVal-----LysAsnIle 966
      370 AGAGTAGCAGCAGTGAAGATGCCAGTGAATGATGATGATGATGATGATGATGATGATGAT 429
      967 AsnValArgArgGlnLysLeuLysGlnHisAlaIleValSerGluLysLeuLeuSerLeu 986
      430 CTGTCCAGAAAGGAGTGAAGCCCAAGCCCAAG-----CAGCCAAAGCTCTCTCTTAAGA 483
      987 LeuProGluTyrValVal-----ProTyrThrIleHisLeuLeuAla 1000
      484 AGGCCAAGAGCTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTT 543
      1001 HisAspProAspTyr-----ValLysValGlnAspIleLeuInLysAsp 1016
      544 AGAAG-CCAAAGATTAACCTGTGACATTAAGCTCAAGATTAAGCCCTCCCAAGCA 602
      1017 ValLysGlnCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsnGlnAsnSer 1036
      603 GCTCGAGCAGCA-----CTAATAATATCCCAATGGTAA 635
      1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
      636 GCAGCCAGTAGCAGCAGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 695
      1057 AspAspAlaLysMetAsnGlnLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
      696 AAGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 755
      1077 MetSerLysSerThrThrTyrSerLeuGlnSerProLysAspProValLeuProAlaArg 1096

```



```
Db      756 CCTAAAAGCAAGTTGTG-----GCCAAGGCCCCAGTGAAGACAGCTACC      800
Qy      1097 Phehehr-----
Db      801 ACCCCTACCCGGAAGAGTTCTAGCAGTGAGGATTCCTCCAGTACGAGAGGAAGAGACAA      860
Qy      1100 ---GlnProAspLysAsnPheserAsnThrLysAsnTyrLeuProProGluMetLysSer      1118
Db      861 AAAAACCACGATGAATAATTAACCAAGTCCCTACAGTTCAGTCCCGG-----      908
Qy      1119 PhehehrProGluLysProLysThrThrAsnValLeuGluAlaValAsnLysProLeu      1138
Db      909 -----CCTTCTGCTCCCGCCACCAAGAGTCTCTGGA---ACCCAGCCTCCCAAG      956
Qy      1139 SerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAla      1158
Db      957 AAGCTGTGGAGAAAGACAGCACCCTGTGAAAGCACT-----GAAAGACAGCATGATGAG      1010
Qy      1159 SerSerSerSerAsn-----
Db      1011 TCTGATTCAGTCTCTGAAGAAGAGAAACCCCACTAAGCAGTAGTCTCTAAGCA      1070
Qy      1164 -----ProSerSerProGluArgLysGlyArgLeuAspSerSerGluMet      1179
Db      1071 ACCACTAAACCACTCCAGCAAGAAAGACAGACAGACCTCTCAGACAGCTCAGACTCT      1130
Qy      1180 AspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGluLysSerAsp      1199
Db      1131 GACAGCTCTGAG---GATGATGAAGCTCTCTTAAGCCAGCTGTACCACCAAGATTTCT      1187
Qy      1200 LysArgAspAspSerSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys      1219
Db      1188 TCAATATAGCCAGCTGTACACCAAGTCACTGAGTGAAGCCAGCTGCAAGCCCGCAAG      1247
Qy      1220 ThrProVal-----ThrGlu      1224
Db      1248 CAACCTGTGGGGGTGGCCAGAACCTTCTGACGAGAAAGCTGACAGCAGCTCCAGTGAG      1307
Qy      1225 GlnGlnGluLysLeuGlyMetAspAspLeuThr---LysLeuValGlnGlnLysPro      1243
Db      1308 GAAGAGAGCAGCTCCAGTGAGAGAGAGAGACAAAGAAAGATGTGGCCACCACTAAGCCC      1367
Qy      1244 LysGlySerGlnArgSer-----ArgLysArgGlyHisThr      1255
Db      1368 AAGCGACTGCCAAGACGCTCTACTCTGCTGCCGCAAGCAGGCTCTCAGGTAAGTAGG      1427
Qy      1256 AlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAsp1Leu      1275
Db      1428 GACAGCAGCTCTGAT-----TCAGACAGCTCC      1454
Qy      1276 GluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys      1295
Db      1455 AGCAGTGAAGAGAGAGAAAGAAACATCTAAGTGTGAGTTAAGAAAGAACCCACAGAG      1514
Qy      1296 ProLeuGlyGlyLysThrPro-----
Db      1515 GTACAGAGAGGTGACAGCCCTTCCAAAGCCAGCCTCTGCAAGAAAGAAAGAGCTGAGAGC      1574
Qy      1303 -----LysGluGluProThrMetLysThrSerLysLys      1313
Db      1575 AGCAACAGTTCTTCTGTGATGACTCCAGTGAGAGAGAGAAAGAAAGCTCAAGGGCAAG      1634
Qy      1314 GlySerLysLysLysSerGlyProProAla-----
Db      1635 GGCTCTCCAGAGACCAAGACCCCAAGGCCAATGGCACCTTGCACTGACTGCCAGAAAT      1694
Qy      1324 -----ProGluGlnGluGlnGluGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGln      1337
Db      1695 GGAAGAGCAGCTAAGAAACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG      1754
Qy      1338 GluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlu      1357
```

```
Db      1755 GTTTCAAATCAGGTTCTATTAAAGACGGAAGCAGAGATGAGCTGCCAAGAGAGCAGAG      1814
Qy      1358 SerProGluSerSerAla1leGlu---SerThrGlnSerThrProGlnLysGlyArgGly      1376
Db      1815 ACTCTCAGGCCAAGAAAGATTAAGCTTCAGACCCCTTAACACATTTCCAAAAAGAGAGAAA      1874
Qy      1377 ArgProSerLysThrProSerPro      1384
Db      1875 GGAGAAAAAAGGCGCATCATCCCA      1898
```

Search completed: September 25, 2003, 04:38:19
Job time : 1030 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:26:40 ; Search time 167 Seconds
(without alignments)
3676.431 Million cell updates/sec

Title: US-09-512-581B-2
Perfect score: 7193
Sequence: 1 MAHSKTRNDGKITYPGCVK.....QKGRGSPSKTPSPQPKKNV 1391

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO/spool/US09512581/runat_23092003.163547.5508/app.query.fasta_1.1543
-DB=Issued_Patents_NA -QFWT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09512581.ecgn1.1.85.6runat.23092003.163547.5508 -NCP=6 -ICPD=3
-NO_MMAP -LARGEOUTERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA:*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PCRTUS.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	349	4.9	530	3	US-09-328-111-682
2	231	3.2	8503	4	US-09-620-312D-130
3	229	3.2	8257	4	US-09-595-684B-30
4	207	2.9	6755	1	US-08-931-999-4
5	201.5	2.8	10136	1	US-08-353-700-2
6	201.5	2.8	10136	1	PCR-US95-16216-2
7	197	2.7	8789	1	US-08-328-254-5
8	191	2.7	10320	4	US-09-091-501B-9
9	190	2.6	30549	4	US-09-134-001C-322
10	189.5	2.6	4868	1	US-08-139-937-12
11	189.5	2.6	4868	5	PCR-US93-11310-12
c 12	184.5	2.6	19307	3	US-08-836-022A-10

c 13	184.5	2.6	19307	3	US-09-427-048A-10	Sequence 10, Appl
14	182.5	2.5	6306	1	US-08-195-487-3	Sequence 3, Appl
15	182.5	2.5	6306	5	PCR-US93-06160-3	Sequence 3, Appl
16	181	2.5	6605	1	US-08-769-309A-4	Sequence 4, Appl
17	181	2.5	6605	1	US-08-994-570-4	Sequence 4, Appl
18	181	2.5	6608	4	US-09-220-132-58	Sequence 58, Appl
19	180.5	2.5	13977	4	US-09-484-970B-60	Sequence 60, Appl
20	180	2.5	8351	1	US-08-198-446B-14	Sequence 14, Appl
21	180	2.5	8351	2	US-08-870-693-14	Sequence 14, Appl
22	179.5	2.5	6306	1	US-08-466-390-3	Sequence 3, Appl
23	179.5	2.5	6306	1	US-08-470-950-3	Sequence 3, Appl
24	179.5	2.5	6306	2	US-08-467-781-3	Sequence 3, Appl
25	179.5	2.5	6306	2	US-08-483-924-3	Sequence 3, Appl
26	179.5	2.5	7453	4	US-09-620-312D-248	Sequence 248, App
27	178	2.5	6152	3	US-08-973-462-1	Sequence 1, Appl
28	177	2.5	6921	4	US-09-643-597-117	Sequence 117, App
29	177	2.5	6921	4	US-09-480-884A-117	Sequence 117, App
30	177	2.5	6921	4	US-09-542-615A-117	Sequence 117, App
31	177	2.5	6921	4	US-09-606-421B-117	Sequence 117, App
32	176.5	2.5	7501	4	US-09-620-312D-249	Sequence 249, App
33	176	2.4	4716	2	US-08-290-731C-1	Sequence 1, Appl
34	175.5	2.4	6775	4	US-09-620-312D-289	Sequence 289, App
35	174.5	2.4	6519	1	US-08-586-985-1	Sequence 1, Appl
36	174.5	2.4	6519	1	US-08-971-988-1	Sequence 1, Appl
37	174	2.4	5361	3	US-08-973-462-2	Sequence 2, Appl
38	173.5	2.4	6773	4	US-09-166-350-27	Sequence 27, Appl
39	171	2.4	580073	4	US-08-545-528D-1	Sequence 21, Appl
40	170	2.4	9370	1	US-08-320-558-27	Sequence 27, Appl
41	170	2.4	9370	5	US-08-545-860D-27	Sequence 27, Appl
42	170	2.4	9370	5	PCR-US94-04496-27	Sequence 25, Appl
43	170	2.4	9391	3	US-08-320-559-25	Sequence 25, Appl
44	170	2.4	9391	3	US-08-545-860D-25	Sequence 25, Appl
45	170	2.4	9391	5	PCR-US94-04496-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-682/C
Sequence 682, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endeigo, Wilson O.
APPLICANT: Steilmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 682
LENGTH: 530
TYPE: DNA
ORGANISM: Homo sapiens
US-09-328-111-682
Alignment Scores:
Pred. No.: 1.94e-24
Score: 349.00
Percent Similarity: 98.53%
Length: 530
Matches: 67
Conservative: 0

Best Local Similarity: 98.53% Mismatches: 1
 Query Match: 4.85% Indels: 0
 DB: 3 Gaps: 0

US-09-512-581b-2 (1-1391) x US-09-328-111-682 (1-530)

QY 37 MetValValIysThrPheMetAspMetAspGlnAspSerGluGluValArgArgLeuTyr 56
 Db 213 ATGGTTGTGAAGAACTTTATGATATGACACAGACTGTGAGAGAGAAAGAGCTTAT 154
 QY 57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspVal 76
 Db 153 TTAACCTACCTTACATCTTGCTGCAGATTTTCTCAAGCATCTGATAAGAGATGT 94
 QY 77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
 Db 93 CGCTTACTGGTAGGCTCTGCTGCTGATATTTTCAGATTTATGCTCTCGAAGGCTCT 34
 QY 97 TyrThrSerProAspLysLeuLys 104
 Db 33 TACACATCCCTGTGATAAATAAG 10

RESULT 2

US-09-620-312D-130
 ; Sequence 130, Application US/09620312D
 ; Patent No. 6569662
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Xue, AIdong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yungqin
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Drmanec, Radoje T.
 ; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
 ; FILE REFERENCE: 784CJP2B
 ; CURRENT APPLICATION NUMBER: US/09/620,312D
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pc_fl_genes Version 1.0
 ; SEQ ID NO 130
 ; LENGTH: 8503
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (91)..(8082)
 ; US-09-620-312D-130

Alignment Scores:

Pred. No.: 6.67e-11 Length: 8503
 Score: 231.00 Matches: 285
 Percent Similarity: 35.41% Conservative: 267
 Best Local Similarity: 18.28% Mismatches: 541
 Query Match: 3.21% Indels: 467
 DB: 4 Gaps: 67

US-09-512-581b-2 (1-1391) x US-09-620-312D-130 (1-8503)

QY 17 ProGlyValIysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
 Db 1087 CCTATGTATATAGGATCA-----ACTGATGAAGCTCTCTCGAAGAGATATGA 1137
 QY 37 MetValValIysThrPheMetAspMetAspGlnAspSerGluGlu----- 51
 Db 1138 -----AAGAAATTAATGATCTTAAACAAATTAAGAGAGGTTCTTAGAGACG 1188
 QY 52 -----GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
 Db 1189 CGGGCTCAGCAATGCAAAAAAGACCAATTGGCCCACTTTGGAGAA-----AAGAT 1242
 QY 67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
 Db 1243 TTGCTTCAGAAAGTACGAAATGAGAAATTAAGAACTTAACACGAGATGCTGTG----- 1296
 QY 83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
 Db 1297 -----ACCTCTCTTCC----- 1308
 QY 103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
 Db 1309 -----CTCAGCTTGCA-----CAGGAATTAAGGCT 1335
 QY 123 ProGluPheAsnArgTyrPheTyrLeuLeuGluAlaSerGlnLeuLeuAspThrVal 142
 Db 1336 AAAAGAAAGAAAGAGTACTTGTGCTGCTGGCAAAATTAACAAATTAAGAAAGCTCAAC 1395
 QY 143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
 Db 1396 -----TATGAGATCAATTT 1410
 QY 163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
 Db 1411 AATATACCAACAAT--ATAACAACAACACATAGCTTTCTAATATTTATATACGA 1467
 QY 183 SerIle-----IleCysGluGluAspThrValSerGlnLeuLeuAspThrVal 199
 Db 1468 GAATATGATGAATCTGCTGTTCAAGTGCATGATTTTCATGAACCTGTGATATACATTA 1527
 QY 200 Leu--ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
 Db 1528 AGTGAGATGAATGCAATCCAGCAACAACCTTAATTCAGGAGATATAGAA----- 1581
 QY 219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238
 Db 1581 ----- 1581
 QY 239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
 Db 1582 -----AGTGAGTTAACACTGCTGCTGCTGATGATATATCG 1620
 QY 259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
 Db 1621 GTATT-----GACTATGAACAACATGACAGCAAGAAAAAATGAGAA 1665
 QY 279 PheLysLeuLysSerAsnAspAsnGluArgLeuGlnValValIysLeuLeuAlaLys 298
 Db 1666 TTGAATTAAGAAAGAAATGATTTGGATGAATTTGAGGCTGTAGAGAAAGAAACTAAA 1725
 QY 299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyr 318
 Db 1726 -----AAGATCAAGAGATGCA----- 1743
 QY 319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
 Db 1743 ----- 1743
 QY 339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
 Db 1744 -----CTAATTCATGAATTTGCAACTTAAGAAATTTAGTAAAGATGAGATATAT 1797

OY	83	CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys	102
Db	1297	-----ACCTCTTCTTCC-----	1308
OY	103	LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuGlnAspThrLysSer	122
Db	1309	-----CTCAGCTTGGAA-----CAGGAATTAAAGGCT	1333
OY	123	ProGlnPheAsnArgTyrPheTyrLeuLeuGlnAsnIleAlaTrpValLysSerLysAsn	142
Db	1336	AAAGAAACAGCAAGGATTACTTGCGCTTGCCAAATTAACAAATAGAACACTCAAC	1399
OY	143	IleCysPheGluLeuGlnAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe	162
Db	1396	-----TATGACGATCAATT	1410
OY	163	SerValIleAsnAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeuMetSer	182
Db	1411	AATATACCAACAAT--ATAACAACAAAAACACTAAGCTTTCTATGAATTTATTATACGA	1467
OY	183	SerIle-----IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal	199
Db	1468	GAATTTGATGATCTGTCTGTTCAGAGCTCTGATGTTTTCAGTAACACSTTGTATACATTA	1522
OY	200	Leu--ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla	218
Db	1528	AGTGAGATAGAAATGCCAGACCAACAAGCTACTAATATCAGGACAAATATGAA-----	1583
OY	219	LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn	238
Db	1561	-----	1581
OY	239	GlnValLeuMetLeuGlyLysThrSerIleSerAspLysSerGlnHisValPheAspLeu	258
Db	1582	-----AGCGAGTTGAACCTACTCTCGTGCAGCATATGATATCTG	1620
OY	259	IleLeuGlnLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu	278
Db	1621	GTATTA-----GACTATGACAACTACGACAGAAAGAAAGAAATGGA	1665
OY	279	PheLysLeuLysSerAsnAspAsnGlnGluArgLeuGlnValValLysLeuLeuAlaLys	298
Db	1666	TTGGAATTTAAAGAAAGAAATGATTTGGATCAATTTGGCTCTAGAAAGAAAACTAA	1725
OY	299	MetPheGlyAlaLysLysSerSerGlnLeuLysSerGlnAsnLysProLeuThrGlnCysTyr	318
Db	1726	-----AAGATCAAGAGATGCA-----	1743
OY	319	LeuGluArgPheAsnAspIleHisValProIleArgLeuGlnCysValLysPheAlaSer	338
Db	1743	-----	1743
OY	339	HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg	358
Db	1744	-----CTAATTCATGAAATTTCCAGACTTAAGAATTAATTAAGTAACATCAGAAATATAT	1797
OY	359	SerHisAspProGluGlnAlaIleArgHisAspValIleValSerIleValThrAlaAla	378
Db	1798	AATCAAGATCTTGGAGATGAATC-----AGTTCA	1827
OY	379	LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu	398
Db	1828	AAAGTAGAGCTGCTTAGAGAAAGACAGACCAAGTTAAGAGACAGAGATACATGAC	1887
OY	399	AspLysArgTrpArgValArgLysGlnAlaMetCkeTglYLeuAlaGlnIleTyrLysLys	418
Db	1888	TCTCAAAAG-----CTAGAAATATTAATAATGCACTTGCTGA-----	1923
OY	419	TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp	438
Db	1924	TATCATTTGGAAAGCATTAAGAACCCAAACAAATGAATGACAGACTCTGTGTGATGCTGAA	1963
OY	439	LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGlnArgIle	458

Db	1984	ACTGTA-----:	2007
Qy	459	PheAlaGlnTyrMetValProHisAsnLeuGluThrTrpGluArgMetLysCysLeu				478
Db	2008	CAATAGCCCTTCTTGAAGATGAAATATGCGCTTGAAGAGAAATATGAAGACTTCA				2067
Qy	479	TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys				498
Db	2068	ACTACATTCACAAATGGA-----				2106
Qy	499	GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProLysThr				518
Db	2107	CAAAAGCCCAATTCGAGCGCAAAAGAAATGCAAGTTGATCTGGAGAAAGAA-----				2157
Qy	519	AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg-----				534
Db	2158	-----TTACAAATTCGCTTTTATATGAGTAACAAACATCCATCCCTTATAGATGGC				2208
Qy	535	-----AsnLeuProAspProGlyLysAlaGlnAspPheMet				546
Db	2209	AAAGTCCAAAGATTTCTCTGTATTTTGGATTTGGAAGAAAGATTTACTGATCTTCAG				2268
Qy	547	LysLysPheThrGlnValLeuGluAspArgLysLysIleArgLysGlnLeuGluValLeu				566
Db	2269	AAAGAACTAAATTAAGAGATGGAAGAAATGAAAGCTTTCGGGAAAGAGATC---ATTTCG				2325
Qy	567	ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys				586
Db	2336	CTTTTCA-----GAATTGAAATCTTTTACCTTCTGGAAGAAAGGCTTAGGAAAGAG				2376
Qy	587	LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu				606
Db	2377	ATA-----CAGACAAATCTGAAAGAGCTCCATATATATACATCGAAAGAAAT				2424
Qy	607	ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal				625
Db	2425	AAATGCTTTCTGAGATGATTCATAAAGAGAGTAGAGTTCAAGGTTTACTTGAAGAAAT				2484
Qy	626	AsnLysSerIleAspGlyThrAlaAspArgLysGlnGlyValProThrAspGlnAla				645
Db	2485	GGGAAACAAAGATGACCTTAGCACTACACAGTCGAATTTAAAGACATGATCAAGAA				2544
Qy	646	Ile-----				646
Db	2545	TTCCAAATTTCCAAACCCCTTCAATGTGACTTTGAGCAAAAGTATAGATGTCCTTAG				2604
Qy	647	-----ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer				663
Db	2605	GAGATGAGAGAAATCAATCAAGAAATAGTTAATCTCTTA-----				2648
Qy	664	PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp-----				680
Db	2647	-----GAAGCCCAAAATTTGATTCGAGTTTGGCTGCTTTGAAGACCGACTTCTTAC				2700
Qy	681	-----AspGluLysValAlaGluAlaIleLeuGlnIlePheLysAsnThr				695
Db	2701	AAGACCCAGAACTTCAGAGAAAAACACGTAGGTTCAAGAAAGCTA-----AATGAG				2754
Qy	696	GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis				715
Db	2755	ATGCAACAGCTGAAGAACAAATTTGAAATATAGAGATCTCCGTCGCAAACTGTGAAGG				2814
Qy	716	HisLysSer-----				718
Db	2815	GAGAAACATCGATTACTGAGAAACCTGCACCAAACTTTGAAGAAGATAAAAACTTTAACT				2874
Qy	719	--LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe				737
Db	2875	CAAGAAAGATGATCTAAACCACTCCAAAGACTTGCAAAT-----				2919
Qy	738	SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro				757

Db 2920 ----GAGAGGACCAACTCAAAAGTGAATTACACGATCTGTTAACTGAATATAGACT 2976
 QY SerAsn-----LeuGlnHisLeuIleThrProLeuValThrIleGlnHis 772
 Db 2977 CAGAACAAATTAGGAATGCTCTTGACTCTGAAACACATCAGAAACAAATTATAC 3036
 QY 773 IleAlaLeuLeuAlaProAspGlnPheAlaIaIaProTyrPheSerTyrPheAlaThrPhe 792
 Db 3037 CTA-----GAGGAAATACAGAGAAACTTAAGATCAATTTGCG----- 3117
 QY 793 IleValLysAspLeuMetLysAsnAspArgLeuProGlyLysThrThrLysLeuTyr 812
 Db 3061 GTTTCACGAAATTTGCAATG----- 3081
 QY 813 ValProAspGlnGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
 Db 3082 -----GAGGAAATACAGAGAAACTTAAGATCAATTTGCG----- 3117
 QY 833 ValArgTyrPheLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
 Db 3118 ----CAAAAGATGGTTGGCATAGATMAAAACAGAGATTGGAAAGCTAAAAATACCC---CAA 3171
 QY 853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLys 872
 Db 3172 ACACCTAACTGCACAGTGTAAAGATATAGAGATATAGACACACAAAGAGATATTTTCT 3231
 QY 873 -----ProAspMetSerArgLeuArgLeuAlaIaIaGlySerAlaIaIaValLysLeuAla 890
 Db 3232 TTATATACAGAGAAATAATGAATGCACAAATGTTAGAGTGTATTAGCAGAAAGAA 3291
 QY 891 Gln-----GluProCysTyrHisGluIleIleThrLeuGlnGlnTyr 904
 Db 3292 CAATTGAGAGACTGACCTAAAGGAAATATGAAATGACCATTTGAAACCCAGAGAAATTA 3351
 QY 905 GlnLeuCysAlaLeuAlaIaIaAsnAspGlnCysTyrGlnValArgGlnValPheAlaGln 924
 Db 3352 AGACCT-----CTTGGGATGAACCTTAATAAGACACAGAGATGTTGCAAA 3399
 QY 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIaIaLecysAla 944
 Db 3400 GAAAGAGACCATGCCATTAAGAAA-----GAAGAGAGCTTTGTAGACCTGTGCAC 3450
 QY 945 LeuCysAla-----LysAspProValLysGluArgAlaHisAlaArgGlnCysLeu 962
 Db 3451 AGACTGCAGAGATTGAAGAAAACTTAAGAAAAAGCCAGCAACTCCAGAAAAAACAG 3510
 QY 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIaIaValSerGlu--- 981
 Db 3511 CAACAACCTCTTAATGTACAGAGAGATGATGATGCCAGAAAAAGATTATAGAAATA 3570
 QY 982 -----LysLeuLeuSerLeuLeuProGluTyrValProTyr 994
 Db 3571 GAGAAATTTAAAGATGAATTAAGAACAAAGAAATTCATTCATTCAGTAAACAGAG 3630
 QY 995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
 Db 3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAATAATATATAGAGAAATGATATACCAA 3690
 QY 1013 GlnLeuLysAspValLysGlnCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsn 1032
 Db 3691 GAAAGAAAAGTCTCAAAGAA-----TTACAGAAAGTCAATTTGAAACAGAG 3735
 QY 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGlnAsnIleLysGlnThrLysAsp 1052
 Db 3736 AGAGACACACCTTAGAGATATATTAAGAAATTCAGACTACAGCCATCAACAAAGAA 3795
 QY 1053 AlaGlnGlyProAspAlaLysMetAsnGlnLysLeuTyrThrValCysAspValAla 1072
 Db 3796 GAACTAAAAAATGGCTCATTTACACTAAAGAACACCAAGAACTATGATGAAGTACA 3855
 QY 1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
 Db 3856 AGAAGCGTA-----TCTGAGAGACACA----- 3876

QY 1093 LeuProAlaArgPheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
 Db 3877 -----GCTCAATATATTAATTAATCTCAGAGACTTGAAGAAAATCCCATACCAAA-----TTA 3924
 QY 1113 ProGluLysLeuLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
 Db 3925 CAGAAAGAGATCCCA-----GTCCTTCAT 3948
 QY 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer---Ser 1150
 Db 3949 GAGCAGACAGAGTTACTGCTTAATGTGAAAAAGTCAAGTCAAGCTCAGCAACAAATGAA 4008
 QY 1151 ArgMetGluThrValSerAsnAlaSerSerSerAsnProSerSerProGlyArgIle 1170
 Db 4009 GAACGTGAGTTATTAACAGAACAGTCCACAAACAGACTCACAACACTGCAGCAAGATTA 4068
 QY 1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
 Db 4069 -----GAAATGGAAGGCTCAGTTGAATGAAAAATTTCAAGAA 4107
 QY 1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
 Db 4108 AGTCAGAGACAGATAAATCTCTTACCAAGAAAGACACACTTAACAGATTAAGAA 4167
 QY 1210 GlnLeuGlnLysProArgGlyArgLysLysThrProValThrGlu----- 1224
 Db 4168 GCCCTTGAAGTTAAACATGACCAAGTGAAGAAACATATTAGAAACCTTGCTAAATC 4227
 QY 1225 -----GlnGluGlnLysLeuGlyMet-----AspAspLeu 1234
 Db 4228 CAGAGCTCTCAAGCAAAACAAAGCAAGTCTCTTAATATGAAAGAAAAAGCAATGAAACT 4287
 QY 1235 ThrLysLeuValGlnGlnGlu-----LysProLys----- 1244
 Db 4288 ACCAAATGCTGAGTGAAGTGAAGCAATTTAAACCCCAAAATTCAGACACTTAAGATA 4347
 QY 1245 -----GlySerGlnArgSerArgLysArgLysHis----- 1254
 Db 4348 GAATATGAATATGCTCGATTTGTCCAAAAGACTTCAAGAAAGTCATATGAAATGAAATCT 4407
 QY 1255 ThrAlaSerGluSerAspGlnGln-----TrpProGluGlnLys 1268
 Db 4408 GTAGCTAAGAGAAAGATGACACAGAGGCTGCAAGAGTCTTCAATCTGAAGAGTAC 4467
 QY 1269 ArgLeuLysGlnAspIle-----LeuGlnAsnGlnAspGlnGln 1281
 Db 4468 CAGCTCAAGAAAAACATTAAGAAATTTAGCTTAACACCTCGAACTGAAAGAGAACTT 4527
 QY 1282 AsnSerProProLysLysGlyLysArgGlyArgProPro---LysProLeuGlnGlyGly 1300
 Db 4528 AAAGTTGCTCATTTGTTCTCGTGAAGAAACAGAGAGAACTATTAAGTTAAGTGAAT 4587
 QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLys-----GlySer 1315
 Db 4588 CTTTCAGAGAGAAAGAACTGAATATCAACATTTCAAAAGGAGTTAAGACAAATCAATGAT 4647
 QY 1316 LysLysLysSerGlyProProAlaProGluGlnGlnGlnGln-----GluArg 1332
 Db 4648 AAATTCAGAAACAAGATCCAGAGATTTATGAGAAAGAGCAACAACCTATATTAACAA 4707
 QY 1333 GlnSerGlyAsnThrGlnGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
 Db 4708 ATTAGTGAGGTTGAGGAAACAGTAATGAACTGAACAAATTCAGAGAGATCGCAAAAGCC 4767
 QY 1353 GlnGlnArgAla---GluSerProGluSerSerAlaIleLeuSerThrGlnSerThrPro 1371
 Db 4768 AAGGATTCACACTACCAAAAGTATTAAGAAAGTAAGATGCTCGAGTT-GACCAAGACTTCA 4826
 QY 1372 GlnLysGlyArg----- 1375
 Db 4827 AGAAAGTCAGAAAGAAATTAACAATTAATGATTAGAGAAAAAGAGAAATGAAAGAGTACA 4886

QY 1376 GlyArgProSerLysThrProSerProSerGlnProLys 1388
||| ||||| |||
Db 4887 GGAGGCCCTTCAGATAGAGAGACCACTGAAAGAAA 4925
RESULT 4
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007
; US-08-931-999-4
Alignment Scores:
Pred. NO.: 9.22e-09 Length: 6755
Score: 207.00 Matches: 78
Percent Similarity: 42.81% Conservative: 50
Best Local Similarity: 26.09% Mismatches: 137
Query Match: 2.88% Indels: 36
Gaps: 9
US-09-512-581b-2 (1-1391) x US-08-931-999-4 (1-6755)
QY 1121 ThrProGlyLysProLys---ThrThrsnValLeuGlyAlaValAsnLysProLeuSer 1139
||| ||||| ||||| |||||
Db 5664 ACAGCACCAAAACCCAGAACGACGACAAAGAAAAAGACAAACCAACCAAGACAA 5723
QY 1140 SerIaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSer 1159
||| ||||| ||||| |||||
Db 5724 GAAAAAGAAAAACAAAGACAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 5783
QY 1160 SerSerSerAsnProSerSerProGlyArgGlyLeuLysGlyArgLeuAspSerSerGlnMet 1179
||| ||||| ||||| |||||

Db 5784 CAGAAAAAGACGACGAGGAAACCGACACACAGGACACAGGACACACACCCAGAAAAACA 5843
QY 1180 AsphHisSerGluAsnGluAspTyrThrMetSerPro---LeuProGlyLysLysSer 1198
||| ||||| ||||| |||||
Db 5844 CCAACACACGACCAACCAAGAAAGACAGACAGCAAAAAACCAACCAAGGAGAGGAAAA 5903
QY 1199 AspLysArgAspAspSerSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLys 1218
||| ||||| ||||| |||||
Db 5904 ACCAAAGAGAAAGACCAAAAAAG-AAAAAGGCACGACGACCAACCAAAAAAGGAAAA 5962
QY 1219 Lys-ThrProValThrGlnGlnGluGlu-----LysLeuGlyMetAspAspLeuThr 1235
||| ||||| ||||| |||||
Db 5963 AAACACGAAACGAAAAAAACAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 6022
QY 1235 LysLeuValGlnGlnGlnLysProLysProLysSerGlnArgSerArgLysArgLysThr 1255
||| ||||| ||||| |||||
Db 6023 AAGAAACAGAAACGACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 6082
QY 1255 ValSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLe 1275
||| ||||| ||||| |||||
Db 6083 GGAGCAGGAAAA-GACCAAGAAAAACACACACGAAACGAAAGCAGGAGA----- 6130
QY 1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys 1295
||| ||||| ||||| |||||
Db 6131 -GAAGGAAACGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 6189
QY 1295 sProLeuGlyGlyGlyThrProLysGlnGluGluProThrMetLysThrSerLysGlyLys 1315
||| ||||| ||||| |||||
Db 6190 GACCAAAAAACAAACGAAAAACGACAAAGAAACAAACAAAGAGAGAAAGAAAGAAAGAA 6249
QY 1315 rLysLys-----LysSerGlyProPr 1322
||| ||||| ||||| |||||
Db 6250 AAAAAACGACGAGAAACAAAGACGAGAAACAAAGAAACCCGCAAGACGAGCC 6309
QY 1322 oAlaProGluGluGluGluGluGluGluArgGluSerGlyAsnThrGlnLys----- 1340
||| ||||| ||||| |||||
Db 6310 CACAAAGAGAGAAAGACCGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 6369
QY 1341 ---SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerPr 1359
||| ||||| ||||| |||||
Db 6370 AACAAAGCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCC 6429
QY 1359 o-----GluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGly---ArgG 1376
||| ||||| ||||| |||||
Db 6430 AAAAAACACACGACGAGGAGAAAGAAAGACCAAAACACACGCAAAAGGCGCAAAAAAGCC 6489
QY 1376 yArgProSerLys-----ThrProSerProSerGlnProLysLys 1389
||| ||||| ||||| |||||
Db 6490 ACGACCAACAAAGAAAGACAAACAGCACAAAGCAAGCAACACACACCCCAAAAAA 6544
RESULT 5
US-08-353-700-2
; Sequence 2, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NOCLETIC ACID ENCODING A
; TITLE OF INVENTION: TRANSLIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORPMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/353,700
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: REED, JANET E.
 REGISTRATION NUMBER: 36,252
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10136 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: HUMAN
 US-08-353-700-2

Alignment Scores:
 Pred. No.: 6,44e-08 Length: 10136
 Score: 201.50 Matches: 283
 Percent Similarity: 35.16% Conservative: 242
 Best Local Similarity: 18.96% Mismatches: 587
 Query Match: 2.80% Indels: 382
 DB: 1 Gaps: 63

US-09-512-581b-2 (1-1391) x US-08-353-700-2 (1-10136)

QY 19 Vallysgluileseraspysile-----Serlysglu 30
 Db 5868 ATGAAGAATTACATCAAAACCTCCATTACAGAGGTACACTAATGACCAAAATTGAA 5927
 QY 31 MetvalargyleuysmetvalvalysThrphenetaspmetaspglinspserglu 50
 Db 5928 GCATGCATAGCATGTAATGGAAAAAATAGTTGGGAACCTTAAGAAAGAAACCTGAGATTTAAGT 5987
 QY 51 GlugluysgluleuTyrr----- 56
 Db 5988 GAAATAATGGAAATATTTTCTTGATCACCAGAGATTACTCCAGAGAGTAGAAACTTCT 6047
 QY 57 -----LeuasnleuAlaleuHisleuAlaSeraspPhepheleuTyrrSHisprogluys 74
 Db 6048 GAAGGCGCTCAATCTGATTTAGAAATGCATGCAGAT-----AAATCATCAGCTGAA 6098
 QY 75 AspvalargyleuValAlaCysCysleuAlaaspIlePheargIleTyrrAlaProglu 94
 Db 6099 GATATTGGAGATATATGGCC----- 6119
 QY 95 AlaProtyrrhrserProaspIysleuysaspIlePhePheleuThrleuThrargIleu 114
 Db 6120 -----AAGTGAATGACACAGCTGGAAGAGAGATTCTTGATGTGGAAAAATGAGCTG 6170
 QY 115 lYsglyleuGlulaspThrlyrSerProglinPheasnargTyrrPheTyrrleuGlulasn 134
 Db 6171 AGTAGATTCAGATCGGAGAAAGCTAGACATTGAGCATGAAGCCCTCTACCTCGAGGCGTGA 6230
 QY 135 lIleAlatrVallysserTyrrasnIleCysPhegluleuGlulaspSerasn-----Glu 152
 Db 6231 TTTAGAGTACTTCAACACAGAGAAAGCTATGTTTAGAAAAAGCAATGAAATTAAGCAGAG 6290
 QY 153 lIlePheThrGlnleuTyrrargThrIlePheSerValIleasnInsglyHisasn----- 170
 Db 6291 GTTATTGTCTGCTTGAAGAAGACTC---TCAGTGTCTACAGTGAAGAAACCCAGCTT 6347
 QY 170 ----- 170
 Db 6348 CGTGGAGATTAGATCTATGTAATAAACAACAGCGCACTGGATCAGTTGTCTGAATAA 6407

QY 171 -----GlnlyValHisMetHisMetValaspIleuMetSerIle--- 184
 Db 6408 ATGAGAGAGAAACACAAAGAGCTTGAGTCTCAAGTCAAGTCTCCATTCATTCAG 6467
 QY 185 lIleCysglulaspThrValIserGln-----GlnleuaspThrValIleuValaspIleu 203
 Db 6468 GTGGCAGAGCGCAGAGGGAAGAAAGACGAACTCTTACAGCTTGTCTGTGATGTG 6527
 QY 204 ValProAlaHislysnleuAsnlysglnAlaTyrraspIleuAlaValaleu----- 221
 Db 6528 -----AGTAGCTGTAAAGAAAGAAACCTATCTCCAGGAAAGCTGCAGAGT 6575
 QY 222 leuysargThrAlaGlnAlaIleGluProtyrrIleThrPhePheasnGlnValleu 241
 Db 6576 TTGGAAAAGACCTCAGCAGCAGCTCTTGACAAATGTGAGCTGGAAGAAACCAATGTCA 6635
 QY 242 MetleuGlylys-----ThSerIleSeraspIleuSerGlnHisVal 255
 Db 6636 CAACCTGATTAAGAGAAAGATTGCTTGACAGGAATCTGAAGCTGCAGGCCAGACTG 6695
 QY 256 PheaspIleuIleuGluleuTyrrasnIleaspSerHisleuIleuSerValleuPro 275
 Db 6696 AGTGAATCAGATTATGAAAGCTGAATGTCTCCAGGCGCTTGAGAGCGCAGCTGTGAG 6755
 QY 276 GlnleuGluPheleuysleuysSeraspaspInsglu-----Argleu 290
 Db 6756 AAAGGTGAGTTCGATGAGAGCTGAGCTCAACACAGGAGAGAGTGCATCAGTGAAGA 6815
 QY 291 GlnValVallysnleuAlaIysMetPheGly-AlaIysaspSerGlnleuAlaSerGln 310
 Db 6816 GGCATCGAAGAACTGAGAGTTCGATTCGAGCCGATGAAAGAAAGAGCTGCACATTCGA 6875
 QY 310 nasnlyspIleuTyrrpInsglyArgPheasnaspIleHisValProIlearg 330
 Db 6876 GAGAAAC-----TGAAGAAGCGGAGCGGAGATG-----ATTCACTTAAGATTA 6923
 QY 330 gleuGluCysVallyspPhealaserHisCys-----Ile 341
 Db 6924 GTTGAGAACCTTGAAAGGAATTCAGATCTCAGAAAGAAACAGAGAGTGTGATCTT 6983
 QY 341 uMetasnHisProaspIleuAlaIysaspIleuThrGlnTyrrleuysValIargSerHisas 361
 Db 6984 GATG-----CCGGAATTCCAAGCGCAGAA-GTAGAGACTCTAAACACAAATAGAGA 7036
 QY 361 ProgluGluAlaIleargHisaspValIleValIserIleValThrAlaAlaIysIysas 381
 Db 7037 GATGGCCAGAAAGCTGAAAGATTTTGAATTAAGCTTGACAGTTAAGCTGTAAGAA 7096
 QY 381 PileuIleuValasn-----AspHisleu 390
 Db 7097 AAATCTGACAAAACAAATACAGAAAAACAGAGTCACTTGACAGACTAGACAAGTTACT 7156
 QY 390 uasnPheValIargGluIargThrleuaspIysArgIrrp-----ArgValIargIysgl 407
 Db 7157 CTCTTCATTAAAGTGTGTAGCAAGAAAGAGCAAGCAGAGATACAGATCAAGAGAGA 7216
 QY 407 uAlaMetMetGlyleuAlaGlnIleTyrrys-----LysTyrrAlaIleuGlnSerAlaI 425
 Db 7217 ATCTAAAACTGAGTGAAGATGCTTCAGATCAGTTAAGAGAGCTAAATGAGCAGAG 7276
 QY 425 aglylyspAlaAlaIysGlnIleAlatrPilelyspIysleuIleuHisIleTyrr 445
 Db 7277 AGCCTGTGTGTGGTCAACCAAAATTAAGAGCGCCAGCAAGAGTCAAGTCAACCCACCA 7336
 QY 445 rGlnasnSerIleaspPargIleuValIargIlePheAlaGlnIlyrMetValPc 465
 Db 7337 AGAGGAAGACATCAGCTGGAATATGACATTGAAAGCTGAGAGCCCGC----- 7385
 QY 465 oHisasnleuGluThrThrGlnIargMetIysCysleuTyrrTyrrleuTyrrAlaThrleuas 485
 Db 7386 -----CTAGAAGCTGATGAAAAAGAGCAG----- 7409
 QY 485 PleuasnAlaVallysnAlaIleuasnGluMetTyrrPlyspCysGlnaspIleuIargHisgl 505

Db	7410	-CTGTGTCTTACACACTGGAAGGAAGTGCACATCATCAATTTACTTAAGGCTAG	7468
OY	505	nValIlyssAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVa	522
Db	7469	AGTGGAGAACTTAAAGAGAGCTAGAGATGATGCCAGGCAAAACCAAGAGCATGCAGCTCT	7528
OY	522	lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl	542
Db	7529	TGAGGCGAGAGAAATTCAAAGGAGAGGTAGAGACCCTAAAGCAAAAATAGACGGATGAC	7588
OY	542	agLAspPhe-----MetLysLysPhe	549
Db	7589	CCAAAGTCCTGAGAGCTCGAATTAGATGTTGTACTATAGTCAGAAAAAGAAATCT	7648
OY	549	eThrGlnValLeuLusAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr	569
Db	7649	GACAAATGAATTACAAAAGAGCAAGACCAATCTGCAATTAACAATTAATAATTCA--	7706
OY	569	oThrCysSerCysLysGlnIleGluGlyCysValArgGluIleThrLysLysLeuGlyAs	589
Db	7707	-----TCATTTGAAATATATTCTTCCAGAAAAAGACAGCAAGAAAGTACA	7750
OY	589	nProLysGlnProThrAsnProPheLeuGluMet-----IleLysPheLeuLe	605
Db	7751	GATCAAAAGAAAATCAACAGCACTGCCATGGAGATCTTCCAAACACAATTAAGAGCTCAA	7810
OY	605	uGluArgIleLeuProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa	625
Db	7811	TGAGAGAGTGCACACCCTCGCATATGACCAAGAACCCGTGAAGCC-----AAAGACA	7864
OY	625	IAsnLysSerIleAspGlyThrAlaAspAspGluAspGluLysValProThrAspGlnAl	645
Db	7865	GAATCTTAGT-----AGTCAAGTAGAGTGTCTTCAACTTGAAGAGGC	7906
OY	645	a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh	664
Db	7907	TCAGTCTCTCAAGAGCCCTTCATGTAGAGCCAAA-----AATATTTATATTGTTT	7954
OY	664	eHisSerAla-----GluThrPheGluSerLeuLeuAlaLysLeuLysMe	679
Db	7955	GCATCTTCAGTAGAAAGGCTCATTCAAAGAGTAGAAGATGCGCAAGCAAACTGGACAA	8014
OY	679	tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysII	699
Db	8015	GAAGAGTAGAAGAAATCAGTAGACAGAAAATCAAT-----	8051
OY	699	eGluGlnAspPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis--	716
Db	8052	---CAAGACCAGAGACAGCTTGTCTCTAACTCTCCAGTGTGGAGAGACCACT	8107
OY	717	---LysSerLysGlyProProArgGlnAlaLysTyrtAlaIleHisCysIleHisAl	735
Db	8108	TTGGAAGAGCAAAACTTAGACATGAGAAATCTACCGTGCATTTGGACAGAAAGATCCA	8167
OY	735	alIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe	755
Db	8168	AGTCTCAATCAAAATAGCGCTCTTGGAGAGACAAATTAGAATGCTGGCAAGATCTCTA	8227
OY	755	uAspProSerAsnLeuHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe	775
Db	8228	C-----AAGATCTTAGAGAT-----GAGCTTGGAAT	8254
OY	775	lLeuAlaProAspGlnPheAlaAlaProTPrLysSerTrpValAlaThrPheIleValLy	795
Db	8255	GACAAAAATGACAAATG-----TCCTTTGTT-----	8282
OY	795	sAspLeuLeuMetAsnAspArgLeuProGlyLysGlyThrThrLysLeuTrpValProAs	815
Db	8283	-----GAAAAGTAAACAAATGATCGCAAGCA	8311
OY	815	pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr	835

Dd	8312	AAC	TGAGCTGCAGAGGCAAAATGCATGAGATGGCACAGAAAACACAGAGCTGCAGAAAGA	8371
Oy	835	pl	euLenglymetLysAsn-----AsnHisserLy	845
Dd	8372	ACT	CAGTCGACAGAAAATAGCTAGCTGGAGAGATTGCAGCTTACTGTTGGAAGAAATAAA	8431
Oy	845	sSer	LyThrSerThrLeuArgLeuThr-----	855
Dd	8432	GAG	CAGCAAGCAAAATCAATTGGAAGGAGCTCCACTGAAAATAGTAATTGAAAGAGCCT	8491
Oy	856	-Thr	lleuLeuSerSerGlyAspLeuThrGlnGlnLysLysLysLysProAspMe	875
Dd	8492	AGATT	GCCTGCGACAAAGAC---CAGGTCGAAAAGGAAGGAAAGG---AAGAGGAAT	8545
Oy	875	tSer	ArgLeuArgLeuAlaIaIaGlySerAlaIaIaValLysLeuAlaGlnGluProCysTy	895
Dd	8546	AGCT	GAATATTCAGCTACGGCTTCATGAAAGCTGAAAAGAA-----	8585
Oy	895	rHis	GluIleIleThrLeuGlnGlnTyGlnLeuCysAlaLeuAlaIleAsnAspGluCy	915
Dd	8586	-CAC	CAGCGCTTGGCTTTGGAC-----ACAAACAAACA	8617
Oy	915	sTy	GlnValArg---GlnValPheAlaGlnLysLeu-----HisLysGlyLeu	931
Dd	8618	GTA	TGAATGAATAAATCCAGACATCCGAGAAATATGACTTTAAGAGAATGCTCTAG	8677
Oy	931	rArg	LeuArgLeuProLeuGlnTyMetAlaIleCys-----AlaIe	945
Dd	8678	TTCA	CAGAGAGCTGGGAGATAGACCTTTTAAATCTAGTAAAGAGAGCTCAATATTCATT	8737
Oy	945	uCys	AlaLysAspProValLysGluArgAlaGlnHisAlaArgGlnCysLeuValLysAs	965
Dd	8738	GAA	AGCTACTCTCAGATTGTTGGAAGAAATGAAAAAACCAAGATGGCAATCTTAATA	8797
Oy	965	nIle	AsnValArgArgGluTyLeuLysGlnHisAlaIaIaValSerGluLysLeuLeuSe	985
Dd	8798	TGT	AAT-----CAGTTGTAACAAAGAAAATGACCTGCCAGGGGAAAATGAA	8845
Oy	985	rLeu	LeuProGluTyValValProTyThrIleHisLeuAlaHisAspProAspTy	1005
Dd	8846	GTT	GTTGTTG-----	8852
Oy	1005	rVal	LysValGlnAspIleGlnLeuLysAspValLysGluCysLeuTrpPheValLe	1025
Dd	8853	-----	ATCAAAATCTCTTAACACAGCTGGAGAGGAAAGAGATGACTGCAGAAAGAACT	8905
Oy	1025	uGlu	IleLeuMetAlaLysAsnGlnAsnAsnSerHisAlaPheIleArg-----	1041
Dd	8906	CTCT	CAACTTTCAACTGCACGAGAAACAGAAAACAGTACTGTATGATTACCAAGT	8965
Oy	1042	---	LysMetValGlnAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaTy	1060
Dd	8966	CGAT	GAATTAACCAACTGAGATCAAAAGAACTAAAGAACTCTTGAAAGAAAAACCAAGGA	9025
Oy	1060	sMet	AsnGluLysLeuTyThrValCysAspValAlaMetAsnIleIleMetSerLysSe	1080
Dd	9026	GCG	AGATGAAATACCTTGATGATGACTGT-----TCCTGCTTATTAACCATGA	9073
Oy	1080	rThr	ThrTySerLeuGlnSerProLysAsp-----ProValLeuProAlaArgPhePh	1098
Dd	9074	A-----	AGTTTGAAGAAAGCTAAAGAAAGATGTAGAGACACACAGAGTCCCTCTGTG	9124
Oy	1098	eThr	GlnProAspLysAsnPheSerAsnThrLysAsnTyLeuProProlumeLtyLysSe	1118
Dd	9125	TTTC	ACAGCAATCTAAACAGATTCGCCGAGGTCCTCTTGCTAGGCTCCAGTTGTT-----	9179
Oy	1118	rPhe	pethThrProGlyLysProLysThrThrAsnValLeuGlnValAlaValAsnLysProLe	1138
Dd	9180	-----	CCAGAGACATCTCCAAATCCCTCTGTACTGAAAAGAGGTT	9220
Oy	1138	uSer	SerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl	1158
Dd	9221	ATCAT	TCTGGCCAAATATATAGCTTCACAGGCAAGGCAAAAGATCCAGTGAATTTGGAGAA	9280

```

QY 1158 aSerSerSerSerSerProSerProGlyArgIleLysGlyArgLeuAspSerSerG1 1178
Db 9281 TGTAGAGAGACCAACACCTGCTACCCA----- 9308
QY 1178 uMetAspHisSerSerLeuAspGlyArgIleMetSerSerProLeuProGlyLysLysSe 1198
Db 9309 -----GAGACCTTTTCT----- 9328
QY 1198 rAspLysArgAspAspSerSerAspLeuValArgSerGluLeuLysProArgGlyArgLy 1218
Db 9329 CAAGAAAGCAGTCTAGTGTATTCACCTGCGAGAA----- 9365
QY 1218 sLysThrProValThrGluGluGluGluLysLeuGlyMetAspAspLeuThrLysLeuVa 1238
Db 9366 -GACACCGAAGGTACTGAGTTTGAGCCAGAG--GACTTCCAGAAAGTTGAAGAAAGG 9421
QY 1238 lGlnGluGlnLysProLysGly-----SerGlnArgSerArgLysArgIleHisThrAl 1256
Db 9422 GTTTGCTGACATCCCGACAGGAAAGACTAGCCATATATCTCGCAGAAACACCATGGC 9481
QY 1256 aSerGluSerAspLysGluGlnGlnTrpProGluGluLysArgLeuLysGluAspLleLeu-- 1275
Db 9482 AACTCGACACGACCCCGCTGGCTGCACAGAAATTCAGCTATCCCGACTGAGTCTCGG 9541
QY 1276 ----GluAsnGluAspLysGlnAsnSerProProLysLysGlyLysArgLysArgProPr 1294
Db 9542 CAAGAAATCTTCTCAGAGTCTCCAAACCAACAGCTGTGGCAGCAGATCACAAAGGT 9601
QY 1294 oLys-----ProLeuGlyGlyGlyThrProLysGluGluProThrMetLys 1309
Db 9602 CAAGTTGCTCAGCGGAGCCAGATGATTCAGGACATCTCCGAGAACCCACACGAA 9661
QY 1309 s-----ThrSerLysLysGlySe 1315
Db 9662 ATCCGTCCAGTCAATATCTCTCTGAGAGAAAGTCCAGTCCAGCCCGACAGAGGGCCT 9721
QY 1315 rLysLysLysSerGly-----ProProAlaProGlu 1325
Db 9722 GAGGCTCAGAGCGGCGGACTTGTCCAGCCCGCAAG 9758

RESULT 6
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-2

Alignment Scores:
Pred. No.: 6,44e-08 Length: 10136
Score: 201.50 Matches: 283
Percent Similarity: 35.15% Mismatches: 242
Best Local Similarity: 18.96% Indels: 382
Query Match: 2.80% Gaps: 63
DB: 5

US-09-512-581b-2 (1-1391) x PCT-US95-16216-2 (1-10136)

QY 19 ValLysGluLysSerAspLysIle-----SerLysGluGln 30
Db 5868 ATGAAAGATATGAGCTCAAAACCTCCATTACGAGGATACAACTAATGACCAAAATTGAA 5927
QY 31 MetValArgArgLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
Db 5928 GCATGCATACATTTGGAAAAATAGTTGGGAACTTAAGAAAGAAACCTCAGATTTAAGT 5967
QY 51 GluLysLysGluLeuLys----- 56
Db 5988 GAAAAATTTGAAATTTTCTTTGTGATCACCAGGAGTTTACCCAGAGTAGAATCTCT 6047
QY 57 -----LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLys 74
Db 6048 GAAGCCCTCAATCTGATTGAAGAAATGATCAGAT-----AATCATCAGCTGAA 6098
QY 75 AspValArgLeuLeuValAlaCysCysLeuAlaAspLysPheArgIleTyrAlaProGlu 94
Db 6099 GATATTGAGATATATGTGGCC----- 6119
QY 95 AlaProTyrThrSerProAspLysLeuLysAspLysPheMetPheIleThrArgGlnLeu 114
Db 6120 -----AAGTGAATGACAGCTGGAAGAGAGATTTCTTGTGGAAGAAATGAGACTG 6170
QY 115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlnAsn 134
Db 6171 AGTAGATCAGATCGGAGAAAGCTAGCATTTGAGCATGAAGCCCTCTACCTGAGGCTGAC 6230
QY 135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn-----Glu 152
Db 6231 TTACAGGTAGTTCAACACGAGACGATATGTTTACAAAAAGCAATGAAATTAACACAGAG 6290
QY 153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn----- 170
Db 6291 GTTATTGTCTGCTTGAAGAGAACTC--TCAGTGTGACAAAGTAGAGAAACCACTT 6347
QY 170 ----- 170
Db 6348 CGTGGAGATTAGATCTATGTCAAAAAACACGCGCAGCTGATCAGTTGCTGAAAAA 6407
QY 171 -----GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
Db 6408 ATGAAAGGAAAAACACAGACCTTGAGTCTCATCAAAAGTGAGTCTCCATTGCATTCAG 6467
QY 185 IleCysGluGlyAspThrValSerGln-----GluLeuLeuAspThrValLeuValAsnLeu 203
Db 6468 GTGGCAGAGGAGAGGAGGAGAAAGAGCAACTCTCTGACATTTGTCTCTCGATGTG 6527
QY 204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu----- 221

```

Db 6528 -----AGTACGCTGTTAAAGACAAACATCTCCAGAAAAAGCTGCAGACT 6575
QY LeuysArgThrIaGlnAlaIleGluProGlyrIleThrPhePheasnGlnValLeu 241
Db 6576 TTGAAAAGAGCTCAGCAGCTGCTTTGACAAAATGAGCTGAAAACCAATTTGCA 6635
QY MetLeuGlyLys-----TherSileSerAspSergLuhHisVal 255
Db 6636 CACCTGATTAAGAGAAAAGATTGCTTGCAGAGATCTGAAAGCTGCAGGCCAGACTG 6695
QY PheAspLeuIleLeuGluLeuTyraSniLeaSerHisLeuLeuLeuSerValLeuPro 275
Db 6696 AGTGAATCAGATTATTAAGAGCTGATCTTCCAGAGCTTGCAGGCCGACATGGTGAG 6755
QY GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu-----ArgLeu 290
Db 6756 AAAGGTGAGTTCGCTTGAAGCTGAGCTCAACACAGAGAGAGTGCATCAGCTGAGAGA 6815
QY GlnValValLysLeuAlaLeuAlaLysMetPheGly-AlaLysAspSergLuhAlaSergL 310
Db 6816 GGCATCGAAGAACTGAGAGTTCGCTTGAAGCCGATGAAGAGAGAGCTGCACATCGCA 6875
QY AsnLysProLeuThrGlnCysTyraLysGlyArgPheAsnAspIleHisValProIleAr 330
Db 6876 GAGAAAC-----TCAGAGAACCGCAGCGGAGAAATG-----ATTCACTTAAGATAA 6923
QY GLeuGluCysValLysPheAlaSerHisCys-----Le 341
Db 6924 GTTGAACCTTGAAGAGAAATTCAGATGTCAGAGAAACACAGAGCTAGTATCTT 6983
QY UMetAsnHisProAspLeuAlaLysAspLeuThrGluTyraLysValArgSerHisAs 361
Db 6984 GATG-----CCGAGAAATTCAGAGCAGAG-CTAGAGACTTAAGAAACCAATTAAGAA 7036
QY ProGluGluAlaIleArgHisAspValIleValSerIleValIleAlaLysLysAs 381
Db 7037 GATGCCAGAGCCCTGAGATTTTTCATTAAGACTTGTACAGTTAAGCTTCGAAAGA 7096
QY PLeuLeuValAsn-----AspHisLeuLe 390
Db 7097 AAATTCGACAAACAAATACAGAAACAAAGAGTCAGTTCAGAGTAAGACAACTTACT 7156
QY uAsnPheValArgGluArgThrLeuAspLysArgTrp-----ArgValArgLysG 407
Db 7157 CTCCTCATTTAAAGCTCTGTAGAGAAAGAGAGCAGACAGATACAGATCAAGAGAGA 7216
QY uAlaMetMetGlyLeuAlaGlnIleTyraLys-----LysTyraLeuGlnSerAla 425
Db 7217 ATCTAAACTGCAGTGCAGATGCTTCAGATCAGTTAAAGAGCTAAATGAGCAGTAGC 7276
QY agLyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyra 445
Db 7277 AGCCTTGTGTGTCAGCAAGAAATTAAGAGCCACAGACAGAGCTAGACCCACCAAT 7336
QY rGlnAsnSerIleAspArgLeuLeuValGluArgIlePheAlaGlnTyraMetValPr 465
Db 7337 AGAGAAAGACATCAGCTGAGAGAAATGCAATTAAGAGCTGAGAGCCGCG----- 7385
QY oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyraTyraIleAlaThrLeuAs 485
Db 7386 -----CTAGAACGTGATGAAGAAAGACAG----- 7409
QY pLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisG 505
Db 7410 -CTCTGCTCTTACAAACACTGMAAGAAAGTGCAGCATGCAATTAATTAAGAGTAG 7468
QY nValLysAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
Db 7469 AGTGAAGAACTTTGAAGAGAGCTGAGATAGCCAGGACAAACCAAGAGCATGCGCTCT 7528
QY lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
Db 7529 TGAGCGACAGAAATTCCAAGAGAGAGTGAAGACCTTAAGAAATTAAGAAAGGATGAC 7588

QY agLysPhe-----MetLysLysPhe 549
Db 7589 CCAAAAGCTGAGAGCTTCGGAATTAAGTGTGTTACTTAAGCTCAGAAAAAGAAATCT 7648
QY eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
Db 7649 GACAAATGATTTACAAAAGAGCAGACAGATATCTGATTAAGAAATTAATTA-- 7706
QY oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuLys 589
Db 7707 -----TCATTTGAAAATTAATTTTCCAGAAAAAGACAAAGAAAGTACA 7750
QY nProLysGlnProThrAsnProPheLeuGluMet-----IleLysPheLeu 605
Db 7751 GATGAAGAAAATTAACACACTGCCATGAGATGCTTCAACACAAATTAAGAGCTCA 7810
QY uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
Db 7811 TGAGAGAGTGGCAGCCCTGCATATGACCAAGAACCTGTAAAGCC-----AAAGACA 7864
QY lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAl 645
Db 7865 GAATCTTAGT-----AGTCAAGTGAAGTGTCTTGAACCTGAGAAAGGC 7906
QY a---lLeuArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
Db 7907 TCAGTTGTACAAAGCCCTTGAAGGCCAA-----AATTAATTAATGTTT 7954
QY eHisSerAla-----GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
Db 7955 GCAATCTTCAGTGAAGAGCCCTCATTCAGAAAGTACAAAGTACAAAGTGCAGAACTGAGA 8014
QY tAspAspLysValAlaAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIl 699
Db 8015 GAAGCATGAGAAATCAGAGTACAGTGAATCAATTAAT----- 8051
QY eGluLysPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis-- 716
Db 8052 ---CAAGACCAAGAGCAGCTGTCTTCAATTCAGAGTCCAGAGTGAAGAGACACCAACT 8107
QY ---LysSerLysLysGlyProProArgGlnAlaLysTyraIleHisCysIleHisAl 735
Db 8108 TTGGAAGAGCAAACTTAAGAACTGAAATGAGAGTGCAGATTTGAGACAGAAATGCA 8167
QY aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
Db 8168 AGTCTACAAATCCAAAAATGCCCTTTCAGAGACACATTAAGAAAGTGCAGAGTCTTA 8227
QY uAspProSerAsnLeuGlnHisLeuIleThrProLeuValIleThrIleGlnHisIleAla 775
Db 8228 C-----AAGAAATTAAGAAAT-----GACCTTGAAAT 8254
QY uLeuAlaProAspGlnPheAlaIleAlaProTrpLysSerTrpValAlaThrPheIleVal 795
Db 8255 GACAAATTCGACAAATG-----TCCTTTGT----- 8282
QY sAspLeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuTrpValProAs 815
Db 8283 -----GAAAAGCTTAACAAAGAAAGTGCAGCAAGAG 8311
QY pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
Db 8312 AACTGAGTGCAGAGAGAAATGCATGAGATGCGACAGAAACAGCAGAGCTCAAGAGA 8371
QY pLeuLeuGlyMetLysAsn-----AsnHisSerLys 845
Db 8372 ACTGATGAGAGAAATTAAGCTAGCTGAGAGTTCGAGTACTGTTGAAAGAAATTA 8431
QY sSerGlyThrSerThrLeuAlaGluLeuThr----- 855
Db 8432 GAGCAGCAAAAGATCAATTTGAAGAGAGCTCAGACACTAGAAATAGTGAATTAAGAAAGCT 8491

FEATURE: CDS
 NAME/KEY: 544...7990
 LOCATION: 544...7990
 US-08-328-254-5

Alignment Scores:

Pred. No.: 1.36e-07 Length: 8789
 Score: 197.00 Matches: 280
 Percent Similarity: 35.77% Conservative: 248
 Best Local Similarity: 18.97% Mismatches: 604
 Query Match: 2.74% Indels: 345
 DB: 1 Gaps: 62

US-09-512-581b-2 (1-1391) x US-08-328-254-5 (1-8789)

QY 19 VallysgluIleSerAspLysIle-----SerLysGluGlu 30
 DB 4057 ATGAAGAATTGAGCTCAAACTCCATTACAGAGGTACAACTAATACCAAAATTGAA 4116
 QY 31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
 DB 4117 GCATGCATGAAATTGGAAAAAATAGTTGGGAACCTTAAGAAAGAAACTCAGATTAAAGT 4176
 QY 51 GluLysGluLysLeuTyr----- 56
 DB 4177 GAAAAATTGGAATATTTCTTGTGATCACCAGAGTTACTCCAGAGTAGAAACTTCT 4236
 QY 57 -----LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLys 74
 DB 4237 GAAGGCCCTCAATTCTGATTGAAATGCAATGCAGAT-----AAATCATCAGGTGA 4287
 QY 75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
 DB 4288 GATATTTGAGATATATGTGCGC----- 4308
 QY 95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 114
 DB 4309 -----AAGGTGAATGACGCTGGAAGAGAGATTTCTTGATGTGAAATGAGCTG 4359
 QY 115 LysGluLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuGluAsn 134
 DB 4360 AGTAGATCAGATCGAGAAAGCTAGCATGAGCTGAAGCCCTTACTCGAGGCTGAC 4419
 QY 135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn-----Glu 152
 DB 4420 TTACAGGTAGTTCAACACGAAAGCTATGTTAGAAAAAGACATGAAATATAGCAGAA 4479
 QY 153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn----- 170
 DB 4480 GTTATTTGTCGCTTGAAGAAGAACTC---TCAGTGTGCACAGAGAGAAACCCAGCTT 4536
 QY 170 ----- 170
 DB 4537 CGTGAGAAATTAGTACTATGTCAAAAAACCAACGCGCATGTGATCACTGTCGAAAAA 4596
 QY 171 -----GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
 DB 4597 ATGAAGAGAAAAACAAAGAGCTTACGTCTCATCAAACTGAGTGTCTCATTCGATTGAG 4656
 QY 185 IleCysGluLysAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
 DB 4657 GTGCGACAGCGCAGGTAAGAAAGAGCAAGCACTCTTCAGACTTGTCTCTCGATGTG 4716
 QY 204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu----- 221
 DB 4717 -----AGTGAGCTGTTAAAGACAAACTCATCTCCAGGAAGAACTCAGAGGT 4764
 QY 222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeu 241
 DB 4765 TTGAAAAAGAGCTACAGGCACTGTCTTGACAAATGTGAGCTGGAAAAACCAATTTGCA 4824
 QY 242 MetLeuGlyLys-----ThrSerIleSerAspLeuSerGlnHisVal 255

DB 4825 CAACTGAATTAAGAAAGAAATTCCTGTGCAAGAATCTGAAGACCTCGAGCCAGACTG 4884
 QY 256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
 DB 4885 AGTGAATCAGATTATGAAAGCTGATGTCTCAAGGCTTGGAGCCGCTGCTGTGGAG 4944
 QY 276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu-----ArgLeu 290
 DB 4945 AAAGGTGAGTTTGCATTAGAGCTGAGCTCAACACAGAGAGAAATGCTACAGCTGAGAGA 5004
 QY 291 GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGln 310
 DB 5005 GGCATCGAAGAACTGAGAGTTGCGCATGAGGCCGATGAAGAAAGAACACCTGCACATCCCA 5064
 QY 310 AsnLysProLeuThrProGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArg 330
 DB 5065 GAGAAAC-----TGAAAGAACCGCAGCGGAGAGATG-----ATTCATTAAAGGTAAA 5112
 QY 330 GLeuGluCysValLysPheAlaSerHisCys-----Le 341
 DB 5113 GTTGGAACCTTGAAAGGGAATTGCAGATGTCAGAGAAACACAGAGCTGATGCTTCTT 5172
 QY 341 MetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
 DB 5173 GATG-----CCGAGAAATCCAAAGCAGAA-GTAGAGCTCTAAAAACAAATGAGAGA 5225
 QY 361 ProGluGluAlaIleArgHisAspValIleValSerIleValIleAlaLysLysAs 381
 DB 5226 GATGCCCGAAGCCGGAAGATTTGTAATGACCTTGACGTTAAGTCTGAAAAAGA 5285
 QY 381 PLeuLeuValAsn-----AspHisLeuLeu 390
 DB 5286 AAATCTGCAAAACAAATACAAAGAAACAAAGTCAGTTGTCAGAACTACAGCAATTACT 5345
 QY 390 AsnPheValArgLeuArgThrLeuAspLysArgTrp-----ArgValArgLysGln 407
 DB 5346 CTCTTCATTAAAGTCTGTAGAAAGAAAGAGACAGACAGATACAGATCAAAAGAGA 5405
 QY 407 ValMetMetLysLeuAlaGlnIleTyrLys-----LysTyrAlaLeuGlnSerAlaAl 425
 DB 5406 ATCTAAACCTCAGTGGAGATGCTTCAGATCAGTTAAAGAGACTAAATGAGCCGATGAC 5465
 QY 425 AgLysAspAlaIleLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTy 445
 DB 5466 AGCCTGTGTGTGACCAAGAAATTTATGAAGCCACACAGAGCTTACAGCCACCAT 5525
 QY 445 GlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
 DB 5526 AGAGGAAGAGCATCAGCTGAGAAATATGATTTGAAAGCTGAGAGCCGCGC----- 5574
 QY 465 HisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
 DB 5575 -----CTAGAACCTGATGAAGAAAGAACAG----- 5598
 QY 485 PLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGln 505
 DB 5599 -CTGTGTCTTTACACAACTGAAGAGAAAGAGAGCATCATGATTAATTAAGGGTAG 5657
 QY 505 nValLysAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
 DB 5658 AGTGAGAAACCTTGAAAGAGCTAGAGATGAGCCAGCAACCAAGAGCATGAGAGCTCT 5717
 QY 522 IleValAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
 DB 5718 TGAGGACAGAGAAATTCAAAGAGAGATGAGAGACCTTAAGAACCAAAATAGAGAGATGAC 5777
 QY 542 agLAspPhe-----MetLysLysPh 549
 DB 5778 CCAAGCTGAGAGCTCGAATTAGATGTTGTTACTATAGTTCAGAAAAAGAAATCT 5837
 QY 549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
 DB 5838 GACAAATGAATTACAAAAAGAGCAAGAGCAATATCTGATTAAGAAATTAATTAATTC-- 5895

OY 569 orthCysSerCysLysGlnIleuGlyCysValArgIleuIleThrLysLysLeuGlyAs 589
 Db 5896 -----TCATTGCAAAATATTGTGCAAGAAAAGACAGCAAGAAATACA 5939
 OY 589 nProLysGlnProThrAsnProPheLeuGluMet-----IleLysPheLeu 605
 Db 5940 GATGAAAGAAAATCAAGACACTCCCATGAGATGCTTCAACACAAATTAAGAGCTCAA 5999
 OY 605 uGluArgIleAlaProValHisIleAspThrIleSerIleSerAlaLeuIleLysGlnVa 625
 Db 6000 TGAAGAGTGGACCCCTGCATATATACCAAGAAAGCTGTAAAGCC-----AAAGCA 6053
 OY 625 IAsnLysSerIleAspGlyThrAlaAspAspGluAspGluIleValProThrAspGlnAl 645
 Db 6054 GAATCTTACT-----ACTCAAGTAGAGTGTCTGCACTGACCTTGAGAGGC 6095
 OY 645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
 Db 6096 TCAGTTGCTTCAAGGCCCTTGATGAGGCCAAA-----AATTAATATATATTGTTT 6143
 OY 664 eHisSerAla-----GluThrPheGluSerLeuLeuAlaCysLysIleSme 679
 Db 6144 GCATCTTCAGTAAATGGCCTCATTCAGAGTAGAGAGATGCCAGCAAGAACTGGAGAA 6203
 OY 679 tAspAspGluLysValAlaGluAlaIleuGlnIlePheLysAsnThrLysSerLysI 699
 Db 6204 GAAGGATGAAGAATCAGTAGACTGAAATCAAAAT----- 6240
 OY 699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis-- 716
 Db 6241 ----CAAGACCAAGAGAGCCTGTCTCTAACTGTCCAGCTGAGAGAGAGACCAACT 6296
 OY 717 ----LysSerLysLysGlyProProArgGlnAlaLysTyraIleHisCysIleHisAl 735
 Db 6297 TTGGAAGAGCAAACTTAGACTGAGAAATCTGCAGCTGGAATGGAGCAAGATCCA 6356
 OY 735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
 Db 6357 AGTGTACAAATCCAAAATGCTCTTTCAGACACACTTAGAAAGTGCTGAGAGTTCTTA 6416
 OY 755 uAspProSerAsnLeuGlnHisIleuIleThrProLeuValIleThrIleGlyHisIleAla 775
 Db 6417 C-----AAGCAATCTAGAGAAAT-----GAGCTTGAAAT 6443
 OY 775 uIleAlaProAspGlnPheAlaIleAlaProTrpLysSerTrpValAlaIlePheIleVal 795
 Db 6444 GACAAAATGACAAAATG-----TCTTTTGT----- 6471
 OY 795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
 Db 6472 -----GAAAAAGTAACAAAATGACTGCAGAAAGA 6500
 OY 815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetIleValArgTr 835
 Db 6501 AACTGTGAGCTGCAGAGGAAATGATGATGAGCAAGAAAACAGCAGAGCTGCAGAGAGA 6560
 OY 835 pIleuLeuGlyMetLysAsn-----AsnHisSerLys 845
 Db 6561 ACTCAGTGGAGAAAATATAGCTAGCTGAGAGAGTTGCACTTGTGGAAGAAAATAAA 6620
 OY 845 sSerGlyThrSerThrLeuArgLeuLeuThr----- 855
 Db 6621 GACACACAAAGATCAATTGAAAGAGCTCACACTAGAAAATATGAAATTAAGAAAGAGCT 6680
 OY 856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
 Db 6681 AGATTTCATGCAAAAAGAC---CAGGTGAAGAGGAAAGGAAAGT---AGAGAGCAAAAT 6734
 OY 875 tSerArgLeuArgLeuAlaIleAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
 Db 6735 AGCTGAATATACACTACGCTTCACTGAAAGCTGAAAAGAAA----- 6774

OY 895 rHisGluIleIleThrLeuGluGlnIleGlnLysCysAlaLeuAlaIleAsnAspGluCys 915
 Db 6775 -CACAGAGCTTGCTGTTGAC-----ACAAACAACA 6806
 OY 915 sTyrglnValArg---GlnValPheAlaGlnLysLeu-----HisLysGlyLeuSe 931
 Db 6807 GTATGAAGTGAATAATCCAGACATACCGAGGAAATGCACTTCAAGAAAGAAATGCTCAG 6866
 OY 931 rArgLeuArgLeuProLeuGlnIleThrMetAlaIleCys-----AlaLe 945
 Db 6867 TTCACAGAACTGGAGATATACCTTTTAAAGTCTAGTAAAGAGAGCTCAATATTCATT 6926
 OY 945 uCysAlaLysAspProValLysGluArgAlaIleHisAlaArgGlnCysLeuValLysAs 965
 Db 6927 GAAAGTACTACACAGATTTTGGAAAGATTAAGAAAACCAAGATGACATCTAAATA 6986
 OY 965 nIleAsnValArgArgGluIleuLysGlnHisAlaIleValSerGluLysLeuLeuSe 985
 Db 6987 TGTAAAT-----CAGTTGAAGAGGAAATGAACGTGCCAGGAGGAAATGAA 7034
 OY 985 rLeuLeuProGluIleTyraValAlaProTyThrIleHisLeuLeuAlaHisAspProAspTy 1005
 Db 7035 GTTGTTC----- 7041
 OY 1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuThrPheValLe 1025
 Db 7042 -----ATCAAAATCCTGTAAACAGCTGGAAGAGGAAAGAGATCTCGCAAGAAAGCT 7094
 OY 1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg----- 1041
 Db 7095 CTCTCAACTTCAAGCTGCACAGAGAGGAGAAACAGTACTGTATATGATACCAAGGT 7154
 OY 1042 ----LysMetValGluAsnIleLysGlnThrIleLysAspAlaGlnGlyProAspAspAlaTy 1060
 Db 7155 CGATGAATTTACACACTGAGATTCAAAGACGTGAAGAAACTCTTGAAAGAAAACCAAGGA 7214
 OY 1060 sMetAsnGluLysLeuTyThrValCysAspValAlaMetAsnIleLeuMetSerLysSe 1080
 Db 7215 GGCAGATGAACTACTGGATAAGTACTGT-----TCTTGCTTATTAAGCATGA 7262
 OY 1080 rThrThrTySerLeuGluSerProLysAsp-----ProValLeuProAlaArgPhePh 1098
 Db 7263 A-----AAGTTAGAGAAGCTTAAGAGATGTAGAGACACAAGATGCATCTGTG 7313
 OY 1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyLeuProProGluMetLysSe 1118
 Db 7314 TTCACAGCAATCTTAACAAACATTCGCCAGAGGCTCTCTTCTAGGTCACATGTT----- 7368
 OY 1118 rPhePheThrProGlyLysProLysThrThrAsnValIleuGlyAlaValAsnLysProLe 1138
 Db 7369 -----CCAGGACCATCTCCAAATCCCTTCTGTACTGAAAAGAGTT 7409
 OY 1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerSnaI 1158
 Db 7410 ATCATCTGGCCAAATAAAGCTTCAGCAAGAGCCAAAGATTCAGTGAATATGAGAGAA 7469
 OY 1158 aSerSerSerSerAsnProSerSerProGlyArgIleTyGlyArgLeuAspSerSergI 1178
 Db 7470 TGGTGAAGAGCAACACACTCTACCCACAGAGACTTTTCTAAAAAAGCAAGAAAGCAGT 7529
 OY 1178 uMetAsp-----HisSerGluAsnGluAspTyThrThrMetSerSerProLeuProG 1195
 Db 7530 CATGAGTGTATTCACCTCTCAGAGAGACAGAGAAAGTACTGAGTTTGAGCAGAGAGACT 7589
 OY 1195 yLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuLysProAr 1215
 Db 7590 TCCAGAAAGTTGTAAGAAAGAGGTTTCTGACATCCGACAGAGAAAGACTAGCCCATATAT 7649
 OY 1215 gGlyArgLysLysThrProValThrGluGlnGlu-----GluLysLeuG 1230
 Db 7650 CTCGCGAAGAACACCATGCACTGGACACAGCCCGGCTGCTGACACAGAAAGTTAGC 7709
 OY 1230 yMetAspAspLeuThr-----LysLeuValGlnGluGlnLysPro--LysG 1245


```

Db      6285 AGCTAGAGAGATTAATCTGTTGGTTAAACAAGAGCTGACGACCTATGCAAAAGAGATCA- 6343
QY      921 aIhPhealaglnlyserhsllysglyleuSerArgleuArgleuProleuGluTyrMetA 941
Db      6344 -----ACCAACGAAATGGAGAAAACCTGCAGAAATTAAGA----- 6379
QY      941 lailecysalaleucysalalysaspProvallysgluArgArgalaHsiAlaArgLnc 961
Db      6380 -----GACCTTAACCTCAAGAAATGGAAGACACTGCTGAAAAA- 6415
QY      961 yslLeuVallylsasnleasnValArgArgGluTyrleuysglNhlsAlaAlaValSerg 981
Db      6416 -----CTCAAAATGGCTGAAT-----AGAACTGAATTTGGAGATG-----CTTTGAG 6455
QY      981 luLysleuLeuSerleuLeuProgluTyrValValProTyrThrIleHslLeuLeuAlaH 1001
Db      6456 ATAAAGTCTGAGTTTA----- 6472
QY      1001 lsaspProaspTyrVallylsValGlnAspIleGluGlnleuLysaspVal----- 1017
Db      6473 -----CTGGAAGAGGATTAATAATTTC-----GAAAGCTTAAGAGCTGAATATGACAT 6521
QY      1018 -----LysGluCysleuTrrPheValLeug 1026
Db      6522 GCAATAGATTTGACAGAGAGTGCCTACCACCTGAGAGGAATGCATC----- 6568
QY      1026 luIlleuMetAlaLysasnGluAsnAsnSerHisAlaHelleArglyMetValGluA 1046
Db      6569 -----CAGAGAGCCAGTT 6581
QY      1046 snlleLysglInThrLysaspAlaGlnGlyProaspAspAlaLysMetAsnGluLysLeuT 1066
Db      6582 CTGTTTCAACAGACAAGATGCTGCTCATCAATGATCCAAAGAGG----- 6628
QY      1066 yrrhValCysaspValAlaMetasnIleIleMetSerlySerThrThrTyrSerLeug 1086
Db      6629 --GTCTAGATATCATCTGCTCAGATATTCCTGCTCAGTCTCATGCTGATCTGGAATTT 6686
QY      1086 luSerProLysaspProvalleuProAlaArgPhePheThrGlnProaspLysAsnPhes 1106
Db      6687 CAATTCTCTGCTGAT-----CTGTATTAACACTATAA 6716
QY      1106 erAsnThrLysAsnTyrLeu-----ProProGluMetLysSer---PhePheThr 1122
Db      6717 CAGAACTAGCCGACGCGTGTATTAATGACACAGATGCTGAAGTCCACATTTGTCACG 6776
QY      1122 roGlyLys-----ProLysThrThrAsnValleuGlyAlaValAsnLysProL 1138
Db      6777 TTGGGAGTGAAGAGATCAATTAAGACCGTTTCCGATGAATAATTACAAAGCGTGA 6836
QY      1138 euSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnA 1158
Db      6837 TAGAAGACGGCCATCTCAGCTGATTAATGTTTATGATGACAGAAATTTGAAAAA 6896
QY      1158 laSerSerSerAsnProSerSerPro-----GlyArgIleLysGly 1173
Db      6897 AAGCTTCAGTTCATATGAGACACAGCATTTACAGAAAAATTGGAAGGCTCAAGAAC 6956
QY      1173 rgleuAspSerSer-----GluMetAspHisSerGluAsnGluAspTyrThrM 1189
Db      6957 AGTGGATGGCACCCAGCATGCGTGGAGCTTAAGACAGCAGCAGCTTGAGCAGATTA 7016
QY      1189 etSerSerProleuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgS 1209
Db      7017 TTGACAGTCTT-----CAGTGGATGACCATTAAGGAGAGAGATGAAGAACTGATGA 7067
QY      1209 erGluLeuGluLysPro-----ArgGlyArgLysThrProValThrGlu 1225
Db      7068 GAAATATAGAGCGCTGACTATATTTTTCACAGCAAGCCGACGCGGATCCACTACCAAA 7127
QY      1235 ln-----GluGluLysLeuGlyMetAsp 1233
Db      7128 AAATTTCTGATTAACCAATACTGTTCAAGAACTGGTCTGAGATGATGATGTCATGG 7187

```

```

QY      1234 -----LeuThrLysleuValGlnGluGlnLysProLysGlySerGlnArgS 1249
Db      7188 CGTTGATTAACGCTCTGCGAAGAACTCTGGAGAAATAT-----GGAGTGAATGACA 7238
QY      1249 erArgLysArgGlyHisThrAlaSerGluSerAspGluGlnInrrProgluGluLysA 1269
Db      7239 CAAGGAAATGTGAAGAAACCC--ACAGAGTACTTAAACATCATG-----ATCA 7286
QY      1269 rgleuysgluaspIleleuGluAsnGluAspGluGlnAsnSerProProLysLys-Gly 1288
Db      7287 ATCTCAACCAAAAGTAT-----GCTGACAGACAGAACCCCTTGGAGCGTGATGGA 7337
QY      1289 LysArgGlyArgProProLysProLysProLysGlyGlyThrProLysGluLupProThrMet 1308
Db      7338 GACGCGTGCAGGCGCTCTCCACAGATC-----TGG 7367
QY      1309 LysThrSerLysLysGlySerLysLysLys 1318
Db      7368 AAACCTCTCTGAAGTGATGCAAGAAAGCAG 7397

```

RESULT 9

```

US-09-134-001C-322
; Sequence 322, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NOCULEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322

```

Alignment Scores:

```

pred. No.: 5 9e-06 30549
Score: 190.00 276
Percent Similarity: 35.29% 235
Best Local Similarity: 19.06% 523
Query Match: 2.64% 414
DB: 4 Gaps: 67

```

US-09-512-581b-2 (1-1391) x US-09-134-001C-322 (1-30549)

```

QY      23 SerAspLysIleSerLysGluGluMetValArgArgleuLysMetValLysThrPhe 42
Db      26830 AATGATGAATATTAATGATTAATCTATGACAAATTCCTAGAAATGCTGAATCAAGA 26889
QY      43 MetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHis 62
Db      26890 TTGTGATTTAAATGAATGACATGAATCACTTAAAGATGAACTTAATATCAAGCTTTTCT 26949
QY      63 LeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCys 82
Db      26950 GTCACAGCAAGCTCAAAATTAATTAATTCAGATGAAGATTTAAACAAACAAATTCACCAT 27009
QY      83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProaspLys 102
Db      27010 GCTTTAAGAAATAGCTGGAAGAAATTCCTGCAAGAAAGAAATGTAATAATTTAGATGAAAA 27069
QY      103 LeuLysaspIlePheMetPheIleThrArgGlnLeuLysGlyLeu-----Glu 118
Db      27070 -----CAAATTCAGGAGCTCAACAAAGATGATGAG 27099

```

```

QY 119 AspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVal 138
   |||||
Db 27100 GATACCTAAAGAT----- 27111

QY 139 LysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyr 158
   :||
Db 27112 ---GCTTTAAATGCTATCCAAAGCTTTATCCAAAGCTTAAGCTAAAGCAATTCATACGCTA 27168

QY 159 ArgThrLeuPheSerValIleAsnAsnGly-----HisAsnGlnLysValHisMetHis 176
   :|||
Db 27169 CAATCTTTA---TCTTATATCAATGATGCACAGCGCTATATGCTGTGAAAAATAAATATTCAC 27225

QY 177 MetValAspLeuMetSerSerIle-----IleCysGluGlyAspThr 190
   |||
Db 27226 AACTCTGATGATTTATCATCTTTAGCAATATACATTATCTTAAGCTAGATTTAGATTAAT 27285

QY 191 ValSerGlnIleLeuLeuAspThrVal-----LeuVal 201
   :|||
Db 27286 GCAATGAAAGCTTACGAGATACATAGAAAGTAATTCACACTTCTGTTCCAAATATAGTGTG 27345

QY 202 AsnLeuValProAlaHisLysAsnLeuAsn-----LysGlnAlaTyrAspLeuAla 218
   |||
Db 27346 AATTATATTATATGCTGATTAAGCAATTTACAATTTGATGAGCGCTACACAAGCA 27405

QY 219 LysAlaLeuLeuLysArgThr----- 225
   |||
Db 27406 AGTCCACAAGCTTCTAAAGCTTCAGAAATCCACAGATTCGAAGAAGTATAGTGTCT 27465

QY 226 AlaGlnAlaIle-----GluProTyrIleThrPhe 236
   :|||
Db 27466 AGTCAGACCATTTACGATACAAAAATATGCTTAATAGTGTACACAGACTTGCACAGAG 27525

QY 237 PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPhe 256
   :|||
Db 27526 AAGAGCAAGATCTAAATTAATTAATAAA---GGATTAAAGATTTAAATTAAGCAACACTT 27582

QY 257 AspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGln 276
   :|||
Db 27583 GAAGATCTCACAAAC-----AAGGTAAATTCACCAAAATCTTTAAACAGATTTATCTTC 27636

QY 277 Leu---GluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGln---ValValLys 294
   |||
Db 27637 CTCCTCAATCAACAGCTTAAGATTTAAACGATTAATAATGAAATATATGAGAGATTAACCTTAA 27696

QY 295 LeuLeuAlaLysMetPheGlyAlaLys-----AspSerGluLeuAla 308
   |||
Db 27697 ACTTTAGTAATCTCTGTTAAAGCAAGTTTAATATAGAAACCGCTATTTAATTTAA 27756

QY 309 SerGln---AsnLysProLeuTyrGlnCys-----TyrLeuGlyArgPheAsnAspIle 325
   |||
Db 27757 CGTCAATTTTAACAAAGCTTTAAAGCAAGCTTAAGCGCTATTTAAATAAATAGCGGTACA 27816

QY 326 HisValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisPro 345
   :|||
Db 27817 AATGTCATATTC-----AATGACATTTCAACT---CTTTTAACCAATA 27858

QY 346 AspLeuAlaLysAsp---LeuThrGluTyrLeuLysValArgSerHisAspProGlnIle 364
   |||
Db 27859 GATATGCTTAAGCAACATTAATGATGTAACGAGCTGTAAAGAACAATCAACAATAATCT 27918

QY 365 AlaIle-----ArgHisAspValIleValSerIleValIlePheAla----- 377
   :|||
Db 27919 GAAGTATTTATTTAATAAGATTAAGATTAATCTTAATAGCTCAAAAAAGCTGCATATAT 27978

QY 378 -----AlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
   |||
Db 27979 AATCAGATTAGAGCGCTTAAGACATTAATAATTAATTAATCAAACTGTTGATTAAGCAATA 28038

QY 394 ArgGluArgThrLeuAspLysArgTyrArgValArgLysGluAlaMetGlyLeuAla 413
   |||
Db 28039 GAA-----TTAATATGATGCTAATGCAAGTTTA--- 28065

QY 414 GlnIleTyrLysLysTyrAlaLeuGln---SerAlaAlaGluLysAspAlaAlaLysGln 432

```

```

Db 28066 -----AAGAACAATGATAGCTCATTTAACAACCACTACAAAAGAC----- 28104

QY 433 IleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArg 452
   :|||
Db 28105 -----AACAATGAATTTAATTAATGCTGATGAAGACCAT 28137

QY 453 LeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGlu----- 469
   |||
Db 28138 AAATTA-----CAATATGATTAACCTTATCAACCTTAAGCGAATATGTT 28179

QY 470 -----ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
   |||
Db 28180 CTTGACAAGAAACGGATACAAATAAAGACCTAATATCATATTTGAAATGATTCAAAAC 28239

QY 484 LeuAsp-----LeuAsnAlaValLysAlaLeuAsnGlu---MetTrpLys 497
   :|||
Db 28240 ATGGATGATGCTAGACACTTCTTAAAGGAATTTGAAAGACTTAAAGATGCTCAAAACAAA 28299

QY 498 CysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLys 517
   |||
Db 28300 GCACATATATGACATTTAAAGATACGCTCAAAACGTCATTTGATGAATTTGAACACGCTAAT 28359

QY 518 ThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPro 537
   :|||
Db 28360 GCAACATCAAAATCTTAAAGCT---CAAGCTAAACAAATGGTA----- 28398

QY 538 AspProGluLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGlu 557
   :|||
Db 28399 -----AATGAG 28404

QY 558 LysIleArgLysGlnLeuGluValLeuValSerProThrCysSer-----CysLysGln 575
   :|||
Db 28405 GAACSTGAAAGAGCGCTTTCTAATATTAATAGCCACATCAATATGATTAGTTAAACAA 28464

QY 576 Ala-----GluGly-----CysValArgGluIleThrLysLys 586
   |||
Db 28465 GCAAAAGATGAAGGGCAATCTGCAATTTGAACACATACATGCAGATGATTAACSTAAAGCA 28524

QY 587 LeuLysAsnProLysGlnProThrAsnProPheLeuGluIleLysPheLeuGlu 606
   :|||
Db 28525 AAACATGATGCTAATCAAAATGATTAACCAAAACCTTAAGATTAATATCTTAATATAGT 28584

QY 607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
   :|||
Db 28585 CAA-----AATCCAACTTATCAATAGAGAAATAAATAAATATATGCTCAAAATTAAT 28638

QY 627 LysSerIleAspGlyThrAlaAspAspGluAspGluLysAlaProThrAspGlnAlaIle 646
   |||
Db 28639 AAGTTAGTAATGGAATTAAGAAATGAATTCACAAACCTTATTAACAAACAATAAGAA 28698

QY 647 ArgAlaGlyLeuGluLeuLeuValLysSerPheThrHisProIleSerPheHisSer 666
   |||
Db 28699 AATGCTCAACAAACATAGATGAAGTCAATGAACCTGTAAAAATTAATTAATGCGCAAA 28758

QY 667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
   |||
Db 28759 GCAGAA----- 28764

QY 687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
   |||
Db 28765 ---GCTAAACAAATGATTAAGAACTTATCAAAAGAAAGACAGAT----- 28806

QY 707 ArgSerAlaLeuLeuProValIleHisHisLysSerLysGluProProArgGlnAla 726
   :|||
Db 28807 -----GCAATTAATTAACACATGATTTTAACACCTTGTCAAAAGGCA 28848

QY 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
   |||
Db 28849 -----CATGCTTTAGCAGATATGATTAACAGAAAAAGATGCACTTCACAT 28896

QY 747 IlePheGluProLeuHisLysSerLeuAspPro-----SerAsnLeuGluHisLeuIle 764
   |||

```

```

Db      28897 ATCGAAAT-----TCTAATTCATTGATGATATCATTAACAATAAGACATGATTT 28950
Qy      765 ThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln----- 780
Db      28951 AAT-----ACTTTAGCTCATATCATTTATTGGGATGACTGATCACAACCATTTAGTT 29001
Qy      781 PheAlaIaPro-----TrpLysSerTrpValAlaThrPheIleVal----- 794
Db      29002 TTTCAGTACCTGAAATTGAGCGCTCAAAATGCTCTAGTAACAAGTGAAGGTGTGTTCAC 29061
Qy      795 -----LysAspLeuLeuMetAsnAspArgLeu 803
Db      29062 AGAGATGAACCTATTTCATTAGATCTATATAAAAAAATGACTTTATGATGAGACTT 29121
Qy      804 ProGlyLysThrThrIleLysLeuTrpValProAsp---GluGluValSerProGluThr 822
Db      29122 -----AAGTCAATATTGTTTCATTACCGAAGACTGATTAAGATGATGATCACCTA 29172
Qy      823 MetValLysIleGlnAlaIle----- 829
Db      29173 ACCGCTAAGTTAGTTATTTAGCTGATGCTCATATGTCACGTAAATGTTCCAGTC 29232
Qy      830 LysMetLeuValAlaTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 849
Db      29233 AAGCTGTAGAAAAAGATTAACAATAGCTAAAAAGGATGCTATATAAAAAACATTTGATGTT 29292
Qy      850 ThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlu---GluGly 868
Db      29293 CTGCTTAACAAAAAATCAAAAGATATGATTTAATACGAATTAACGATTAACCTACATCAACGT 29352
Qy      869 LysIleSerLysProAspMetSerArgLeuArgLeuAlaIaIleValIleValIys 888
Db      29353 GAAGATGCCAAAGCTGAATTAAGAAATGAAAGTTGAAA-----AAGCAAGCCATGATTA 29403
Qy      889 LeuAlaGlnGluProCysTyrHisGluIleIleThr----- 900
Db      29404 GTGATCATTCAAAATGATTAAGATTAAGATTAAGAAACGTAACGATTTTGAAAGAA 29463
Qy      901 LeuGlnGlnIleArgIleLeuCysAlaLeuAlaIleAsn----- 912
Db      29464 ATACATACAGTTTGATCTTAAACGCTTACGCTAATTAAGCTAAAGATATCATTA 29523
Qy      913 AspGlnCysTyrGlnValAlaArgIleValPheAlaGln---LysLeuHisIleGlyLeuSer 931
Db      29524 GATGTTATATCTCAATTCCAAAATGCTTCAAGAAATTAAGAAATTAAGAAATTAAGGTTTAACT 29583
Qy      932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProVal 951
Db      29584 TCTAATGAAAAAACTCAGTTTGATTAACAATTAACCTAAGCTAACAAGAAATTTTAGAA 29643
Qy      952 LysGluArgArgAlaHisIleAlaArgGlnCysLeuValLysAsnIleAsnValAlaArgGlu 971
Db      29644 AAGATCGACATGCTCATTAAT-----TTAGTAGAATTAATTAATTAATTAATTAATTAATTA 29694
Qy      972 Tyr-----LeuLysGlnHisAlaIleAlaIleAspIleLeuLeu 984
Db      29695 TTTAATATATGATATTAACAATTTTAAACCAACGACATTAACGAGTAA----- 29748
Qy      985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuAlaHisAspProAsp 1004
Db      29749 -----CATATAGCAGAAACAATAAATTTAGCA 29772
Qy      1005 TyrValLysValGlnAspIleGlnIleLysAspValLysGlnCysLeuTrpPheVal 1024
Db      29773 TATGTTTATAGTAACAAACTGACGAAATACTAATATTAATCAATGCTTCTTACTTTATA 29832
Qy      1025 LeuGlu----- 1028
Db      29833 AAACAATGGGCACTTGATAGAAATTAACAATTCACAGAAAGATGAATTCATTCGT 29892
Qy      1029 MetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLys 1048
Db      29893 GGTGCGCATACCGTACCAAGATGTACAAAGCATTAATTAACAGGTATAGACAAATCTTG 29952

```

```

Qy      1049 GlnThrLysAspAla----- 1053
Db      29953 AAAGTAATGTACAGTATATTAATCAATCTTCAACGATTCCTGCAATTAATTAATAT 30012
Qy      1054 ---GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
Db      30013 CTTCATTCAAAATTTGATGCTAGATTAAGAGAAAAG-----GATGTGCA 30057
Qy      1073 MetAsnIleIleMetSerLysSerThrThrTyrIleSerLeuGlnSerProLysAspProVal 1092
Db      30058 AACCATTCGTACCAACAGTGAACAACATTCAAAGAGTTCYAAAGAGA-----ACGGGTGT 30111
Qy      1093 LeuProAlaArgPhe-----PheThrGlnProAspLysAsnPheSerAsnThrLysAsn 1110
Db      30112 GAACCAAGTAAATCAACAAAGAAACACGACACCAACCAACTTCATAGACATTAATGAT 30171
Qy      1111 TyrLeuProProGluMetLysSerPhePhe-----ThrProGlyLysProLysThrThr 1128
Db      30172 AGCCTATTCAAAACATTTAGTTGATTAATTTGCGCAAAACGTAGGCTTATTAACATTA 30231
Qy      1129 AsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys 1148
Db      30232 GGTTCATCTTCTAGTCTTCTGCTAGTTTGGCTAAAGACGTAATAAAAAAGAAAGAAAGAA 30291
Qy      1149 SerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly 1168
Db      30292 AAACAATCGATTA----- 30306
Qy      1169 ArgIleLysGlyArgLeuAspSerSerGlnMetAspHisSerGluAsnGluAspTyrThr 1188
Db      30307 -----AATCATCAACAAGATTAATTCGT 30327
Qy      1189 MetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArg 1208
Db      30328 CTTCACAT-----ACTGATTAATATAGATCCAAATTTGATTAACAGTAAG 30369
Qy      1209 SerGlnLeuGlnLysProArgGlyArgLysLysThrProValThrGlnGlnGlnGlnLys 1228
Db      30370 CGTAAATATGATTA-----GAAAGACAA 30393
Qy      1229 LeuGlyMetAspAspLeuThrLysLeu-----ValGlnGlnGlnLysProLysGlySer 1246
Db      30394 ATTCAAAACGATGACAAACATTAATTCAGTGTGCTAACAATTAAGAAATTAAGAAAG 30453
Qy      1247 GlnArgSerArgLysAlaGlyHisThr-----AlaSerGlnSerAsp 1260
Db      30454 CAATTGAGTGAAGAGATATTCATTAATCCCGCTGTTAAGCGTAAACAAACAAACAGTAT 30513
Qy      1261 GluGlnGlnIleProGlnGlnLys 1268
Db      30514 AACCAAGATTAACAACAGAAAGAA 30537

RESULT 10
US-08-139-937-12
; Sequence 12, Application US/08139937
; Patent No. 5821070
;
GENERAL INFORMATION:
; APPLICANT: LEE, WEN-HWA
; APPLICANT: SHAN, BEI
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/139,937
 FILING DATE: 20-NOV-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: CAMPBELL, CATHERYN
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 9370
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-9001
 TELEFAX: 619-535-8949
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4868 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-139-937-12

Alignment Scores:
 Pred. No.: 2,52e-07 Length: 4868
 Score: 189.50 Matches: 277
 Percent Similarity: 35.50% Conservative: 247
 Best Local Similarity: 18.77% Mismatches: 608
 Query Match: 2.63% Indels: 345
 Gaps: 61

US-09-512-581b-2 (1-1391) x US-08-139-937-12 (1-4868)

QY 19 VallysgluIleaserasplysile-----SerlysgluIleu 30
 Db 156 ATGAAGAATTTACACCAAACTCCATTACAGAGAGTACACTAATGACCAAAATTGAA 215
 QY 31 MetvalaIargIleuIysmetvalIalIysthrPhemetaspmetaspIleaspserglu 50
 Db 216 GCATGCATGAAATTCGAAAAAATATGTTGGGGAACCTTAGAAGAAACAACTCAGATTAACT 275
 QY 51 GluGluIysGluIleuIyr----- 56
 Db 276 GAAAAATTCGAAATATTTCTTGTGTATCACACGAGCTTACTCCAGAGAGTAGAACTTCT 335
 QY 57 -----LeuasnleuIaleuIhisleuIalaseraspPheheuleuIhisProgluIys 74
 Db 336 GAAGGCTCATTTCTGATTAGAAATGCATGCAGAT-----AAATCATCAGCTGAA 386
 QY 75 AspvalaIargleuIleuValaIalacysCysleuIalaspIlepheargIleTyralaProglu 94
 Db 387 GATATTGGAATATGATGCGC----- 407
 QY 95 AlaProTythSerProaspIysleuIysaspIlePheMetheIlethraIgluIleu 114
 Db 408 -----AAGTGAATGACACCTGGAAGGAGAGATTCTTGTATGTGAAAAATGAGCTG 458
 QY 115 LysGluIleuGluIysPheIysSerProgluIleuIhisIleuIleuIleuIleu 134
 Db 459 AGTAGATCAGATCGAGAAAGCTAGCATGAGCAAGCCCTTACCTGAGAGCTGAC 518
 QY 135 IlealatrVallysserIyrasnIleCysPhegluIleuIleuIleuIleuIleu 152
 Db 519 TTAGAGTAGTTCACACAGAGAGCTATGTTAGAAAAAGACATAAATAAGCAGAG 578
 QY 153 IlePheThrGluIleuTyraIargIthrIleuPheSerValIleasnIleuIhisasn 170
 Db 579 GTTATTGTCTGCTTGAAGAGAACTC-----TCAGTGTGCACAGTAGAGAAACCACTT 635
 QY 170 ----- 170

Db 636 CGTGGAAATTCATCTATGTCAAAAAACCCAGGCGCTGCATCAGTTCTGTAAAAA 695
 QY 171 -----GluIysValIhisMetIhisMetValaIaspIleuIleuIleuIleuIleu 184
 Db 696 ATGAAGGAGAAAAACAGAGGCTTGATGTCATCAAAAGAGAGTGTCCATTCATTCAG 755
 QY 185 IleCysGluIlyaspIthrValSerGln-----GluIleuIleuIleuIleuIleuIleu 203
 Db 756 GTGGCAGAGGACAGAGGTGAAGGAAAGACGACTCTTCAGACTTGTCTGTGTGTG 815
 QY 204 ValProIalHisIysasnIleuIasnIleuIleuIleuIleuIleuIleuIleuIleu 221
 Db 816 -----AGTAGCTGTAAAAAGCAAAACATCTCCAGGAAAGCTGCAGACT 863
 QY 222 LeuIysArgThrAlaGlnAlaIleGluProTyrlleThrIlePheasnIleuIleu 241
 Db 864 TTGGAAGAGACTCACAGCAGCTGCTTGCACAAATGTGAGCTGGAACCAATTCGA 923
 QY 242 MetleuGluIlys-----ThrSerIleSeraspIleuSerGluIhisVal 255
 Db 924 CAACCTGAATTAAGAGAAATTCCTTGTCAAGGAATCTGAAGCCTGCAGGCCAGACTG 983
 QY 256 PheaspIleuIleuGluIleuTyraIasnIleaspSerIhisleuIleuIleuIleuIleu 275
 Db 984 AGTGAATTCAGATTATGAAAGCTGAATGTCTCCAGGCCCTTGAGGCCGACTGTGAG 1043
 QY 276 GluIleuGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 290
 Db 1044 AAAGGTGATGTCATTCAGAGCTGAGCTCAGACACAGAGAGAGAGTGCATCAGTACAGA 1103
 QY 291 GlnValIalIysleuIleuIalIalIysMetPheGly-AlaIlysaspSerGluIleuIleuIleu 310
 Db 1104 GGCATGAGAACTGAGAGCTGCAATTCAGAGCCGATGAAGAAAGAGAGCTGCACATCGA 1163
 QY 310 nasIlyspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 330
 Db 1164 GAGAAAC-----TGAAGAACCCGAGCGGAGAAAG-----ATTCACCTTAAGTAAAA 1211
 QY 330 GluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 341
 Db 1212 GTTGAGAACCTTGAAAGGGAATTCAGAGATTCAGAAAGAAACCCAGAGCTAGTATCTT 1271
 QY 341 UmetasnIhisProaspIleuIalIalIysaspIleuIleuIleuIleuIleuIleuIleu 361
 Db 1272 GATG-----CCGAGATTCACAAAGCAGAA-GTAGAGACTTAAACAAACAAATAGAGA 1324
 QY 361 ProgluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 381
 Db 1325 GATGGCAGAGGCTGAAAGTTTGAATTAGACCTTGCACAGCTTAAGGTGAGAAAGAA 1384
 QY 381 PileIleuIleuValasn-----AspIhisleuIleu 390
 Db 1385 AAATCTGACAAACAAATACAGAAAGAAACAGGTGCTGTGAGACTGACAACTTACT 1444
 QY 390 UasnPheValaIargIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 407
 Db 1445 CTCTTCATTTAAAGCTGTGTAGAAAGAAAGAGCAGAGAGATCAATCAAAAGAA 1504
 QY 407 UalameIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 425
 Db 1505 ATCTAAACCTGCAGTGCAGTCTTCAGATCACTTAAAGAGACTTAATGAGGAGCTAGC 1564
 QY 425 agIlyaspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 445
 Db 1565 AGCTTGTGTGTGAGCAGAAATTTGAAGGCGCAGAAAGAGTGTGAGCCAGCAAT 1624
 QY 445 rgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 465
 Db 1625 AGAGGAAGAGCATCAGCTGAGAAATAGCATGTGAAAGCTGAGAGCCGCGC----- 1673
 QY 465 ohIasnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 485
 Db 1674 -----CTAGAGCTGATGAAAGAGCAG----- 1697


```

Db      3509 ATCATCTGGCCAAATTAACCTTCAGCAGAGGCAAGATCCAGTGAATATATGCGAGAA 3568
Qy      1158 aserSerSerSerSerpSerProglYargIleYlsgIYArgLeuAspSerSergI 1178
Db      3569 TGGTGGAGGACCAACACCTCTGACCCAGAGACCTTTTCTAAAAACCAAGAAAGCAGT 3628
Qy      1178 uMetAsp-----HisSerGluAsnGluAspTyrThrMetSerSerpLeuProgl 1195
Db      3629 CATGAGTGTATTCCACCTCGACAGAGACGGAAGTACAGTGTAGCGCAGAGGAGCT 3688
Qy      1195 YLysYlSerAspYlArgAspAspSerSerpLeuValArgSergIleuGluYlProAr 1215
Db      3689 TCCAGAAAGTTGTAAGAAAGGTTTGTGACATCCACAGCAAGAAAGCTAGCCATATAT 3748
Qy      1215 gglYArgYlSerYlSerProValThrGluGlu-----GluYlSerGlu 1230
Db      3749 CTGCGAGAGAACACCACTGGCAACTGGAGAGAGCCCGCGCTGCTGCACACAAAGTTACC 3808
Qy      1230 YMetAspAspLeuThr-----LYsLeuValGluGluGluYlPro--LYsGlu 1245
Db      3809 CCTATCCCACTGACTGTCCCAACAAATCTTGAGAGAGTCTCCCAACCAAGAGCTGG 3868
Qy      1245 YSergIleArgSerpArgYlS--ArgGlyHisThrAlaSerGluSerAspGluGluGlnTr 1264
Db      3869 TGGCAGCAGATACCAAAAGGTGAAAGTTGCTCAGCGAGGCCCAAGTAGATTTCAGGCACAT 3928
Qy      1264 pProGluGluYlSerYlSerGluAspIleLeuGluAsnGluAspGluGlnAsn----- 1282
Db      3929 CCTCCGAGAACCCACCAAGAAATCCGTCACAGTCAATATCTTCTTAGAGAAATCCGAC 3988
Qy      1283 -SerProYlSerYlSgIY-----LYsArgGlyArpProYlSerProLeuGlu 1298
Db      3989 TGACAGCCCGCAGAGAGGCGCTGAGAGGTCAAGCGCGCGCACTTCCCGAGCCCAAGC 4048
Qy      1298 YglY-GlyThrProYlSerGluGluProThrMetYlSerYlS 1312
Db      4049 TGGACTGTGAGTCCAAAGGCGAGTGAAGACTGTAAAGTCCAGTGA 4092

RESULT 11
PCT-US93-11310-12
; Sequence 12, Application PC/TUS9311310
; GENERAL INFORMATION:
; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11310
; FILING DATE: 19-NOV-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: FP-CJ 9790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4868 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US93-11310-12

Alignment Scores:
Pred. No.: 2,52e-07 Length: 4868
Score: 189.50 Matches: 277
Percent Similarity: 35.50% Conservative: 247
Best Local Similarity: 18.77% Mismatches: 608
Query Match: 2.63% Indels: 345
Dbs: 5 Gaps: 61

US-09-512-581b-2 (1-1391) x PCT-US93-11310-12 (1-4868)

Qy      19 ValYlSerGluIleSerAspYlSle-----SerYlSerGluGlu 30
Db      156 ATGAAAGAAATTAAGCTCAAACTCCATTACAGAGGTACACATTAATGACCAAAATTTGAA 215
Qy      31 MetValArgYlSerYlSerMetValValYlSerThrPheMetAspMetAspGluAspSerpGlu 50
Db      216 GCATGCATAGATTTGAAAAAATATGTTGGGCACTTAAGAAAGAAACTCAGATTTTAAGT 275
Qy      51 GluGluYlSerGluLeuYlS----- 56
Db      276 GAAAAATTTGAAATTTTCTGTGATCCACAGAGGTACTCCAGAGTAGAAGAACTTCT 335
Qy      57 -----LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuYlSHISProGluYlS 74
Db      336 GAAGCCTCAATTTGATTTAGAAATGATGACAT-----AAATCATCAGCTGAA 386
Qy      75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleYlAlaProGlu 94
Db      387 GATTTGTGAGATTAATGTGGCC----- 407
Qy      95 AlaProYlThrSerProAspYlSerYlSerYlSerPheMetPheIleThrArgGluLeu 114
Db      408 -----AAGGTGAATGACAGCTGAGAGAGAGATTTCTGTGATGGAATAATGACCTG 458
Qy      115 YlSerGluLeuGluAspThrYlSerProGluPheAsnArgYlPheYlThrLeuGluAsn 134
Db      459 AGTAGATTCAGATCGGAGAAAGCTAGCATTCAGCATGAGCCCTCTACCTGAGAGCTGAC 518
Qy      135 IleAlaThrValYlSerYlSerYlSerYlSerYlSerYlSerYlSerYlSerYlSerYlSerYl 152
Db      519 TTAGAGTAGTTCAACAGAGAACTATGTTTACAAAAACATGAAATTAAGCAGAG 578
Qy      153 IlePheThrGluLeuYlSerYlSerYlSerYlSerYlSerYlSerYlSerYlSerYlSerYl 170
Db      579 GTTATTTGCTGCTTGAAGAAAGCTC---TCAAGTGTCAAGTAGAGAAACCAAGCTT 635
Qy      170 ----- 170
Db      636 CGTGGAGAAATTAAGTACTATGTCAAAAAAACACGCGCACTGGATGATGTGTGAAAAA 695
Qy      171 -----GlnYlValHisMetHisMetValAspLeuMetSerSerpIle--- 184
Db      696 ATGAGAGAGAAACACAAAGAGCTTGAGTCTCATCAAGTAGTGTCTCCATTCATTCAG 755
Qy      185 IleCysGluGluYlSerThrValSergIle--GluLeuLeuAspThrValLeuValAsnLeu 203
Db      756 GTGGCAGAGCAGAGGTGAAGGAAAGACGCACTCTTCAGACTTGTGCTCTGAGGTG 815
Qy      204 ValProAlaHisYlSerAsnLeuAsnYlSerGluAlaYlSerValAlaYlSerValAlaLeu 221
Db      816 -----AGTGAAGCTGTTAAAGACAAATCATCTCCAGGAAAGCTGCAGAGT 863
Qy      222 LeuYlSerThrAlaGluAlaIleGluProYlIleThrThrPhePheAsnGluValLeu 241
Db      864 TTGAAAAAGAGCTCACAGGCACTGTTTACAAAAATGTAGCTGGAAGAAACCAATTTGCA 923
Qy      242 MetLeuGluYlS-----ThrSerpLeuSerGluHisVal 255

```

```

Db      924 CAACGAAATGAAGAAAGAAATTCCTTGCACGAATCTGAAGAGCTGAGGCCAGACTG 983
      |||
      |||
      |||
Qy      236 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSerValLeuPro 275
      :
      :
      :
Db      984 AGTGAATCAGATTAATGAAGAGCTGAATCTTCACAGAGCTGAGGCCGACTGGTGGAG 1043
      :
      :
      :
Qy      276 GluLeuGluPheLeuLeuLeuSerAsnAspAsnGluIu-----ArgLeu 290
      :
      :
      :
Db      1044 AAAGGTGAGTTCGCTTGAAGAGCTGAGCTCAGACAGAGAGGAAGTGCATCAGCTGAGAGA 1103
      :
      :
      :
Qy      291 GluValValLeuLeuAlaLeuMetPheGly-AlaIysAspSerGluLeuAlaSerG1 310
      :
      :
      :
Db      1104 GGCATCGAGAAAGCTGAGAGTTCGATGTGAGGCCGATGAAGAAAGACAGCTGCACATCGCA 1163
      :
      :
      :
Qy      310 AsnIlySerProLeuTyrPheGlyTyrLeuGlyArgPheAsnAspIleHisValProIleAr 330
      :
      :
      :
Db      1164 GAGAAAC-----TGAAAGAACCGCGAGCGGAGAAATG-----ATTCACTTAAGGTAAAA 1211
      :
      :
      :
Qy      330 GluGluIuGlyValIysPheAlaSerHisCys-----Le 341
      :
      :
      :
Db      1212 GTTGAAGACCTTGAAGAGGAATTCAGATGTCAGAGAAACAGAGAGCTAGTCTT 1271
      :
      :
      :
Qy      341 UmetAsnHisProAspLeuAlaIysAspLeuThrGluTyrLeuIysValArgSerHisAs 361
      :
      :
      :
Db      1272 GATG-----CCGAGAAATTCAGAAAGCAGAA-CTAGAGAGCTTAAGAAACAAATAGAGA 1324
      :
      :
      :
Qy      361 ProGluGluAlaIleArgHisAspValIleValSerIleValIleAlaIysLysAs 381
      :
      :
      :
Db      1325 GATGCCAGAACCCCTGAAAGTTTGTGAATTAGACCTTGCAGTTAAGCTGAGAAAGA 1384
      :
      :
      :
Qy      381 PileuLeuValAsn-----AspHisLeuLe 390
      :
      :
      :
Db      1385 AAATCTGCAAAACAAATACAGAAAGAAAGAGTTCAGACATAGCAAGACTTACT 1444
      :
      :
      :
Qy      390 uAsnPheValArgGluArgThrLeuAspLysArgTyr-----ArgValArgLysG1 407
      :
      :
      :
Db      1445 CTCCTCATTTAAAGTCTGTAGAGAAAGAGAGCAGACAGATACAGATCAAGAGAGA 1504
      :
      :
      :
Qy      407 uAlaMetMetGlyLeuAlaGlnIleTyrLys-----LysTyrAlaLeuGlnSerAlaI 425
      :
      :
      :
Db      1505 ATCTTAAACCTCAGCGAGATGCTGCAGAAATCAGTTAAAGGAGCTTAATGAGGCGAGTAC 1564
      :
      :
      :
Qy      425 acLYAspAlaIalalysGlnIleAlaTPrIleLysAspLysLeuHisIleTyrTy 445
      :
      :
      :
Db      1565 AGCCTGTGTGTGAGCCAGAAATTAATGAGGCCACAGACAGACTTACAGCCCAAT 1624
      :
      :
      :
Qy      445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
      :
      :
      :
Db      1625 AGAGCAGAGAGCATCAGCTGAGAAATAGCATGAAAGCTGAGAGGCCCGC----- 1673
      :
      :
      :
Qy      465 oHisAsnLeuGluThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
      :
      :
      :
Db      1674 -----CTAGAACCTGATGAAGAAAGACAG----- 1697
      :
      :
      :
Qy      485 PLeuAsnAlaValIysAlaLeuAsnGluMetTyrLysCysGlnAsnLeuLeuArgHisG1 505
      :
      :
      :
Db      1698 -CTGTGTGTCTTACACAGACTGAGAGAAAGTGAAGATGATGCAAGTTTACTTTAAGGATAG 1756
      :
      :
      :
Qy      505 nValIlyAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
      :
      :
      :
Db      1757 AGTGAAGAACTTGAAGAGAGAGCTGAGATAGCCAGCAAAACCAAGAGCATGCGAGCTCT 1816
      :
      :
      :
Qy      522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
      :
      :
      :
Db      1817 TGAGCAGAGAAATTCAGAAAGAGAGAGTGAAGCCTTAAGAGCAAAATAGAGGATGAC 1876
      :
      :
      :
Qy      542 agLAspPhe-----MetLysLysPh 549
      :
      :
      :
Db      1877 CCAAGGTCGAGAGCTCGAATTAATGATGTTTACTTAAGCTAGAGAAAGAAATCT 1936
      :
      :
      :
Qy      549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
      :
      :
      :

```

```

Db      1937 GACAAATGAATTAACAAAAAGACAGAGAGGAATATCTGAATTAAGAAATTAATATCA-- 1994
      :
      :
      :
Qy      569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
      :
      :
      :
Db      1995 -----TCATTTGAAAAAATTTTTCAGCAAGAAAGAGCAAGAGAAAGTCA 2038
      :
      :
      :
Qy      589 nProLysGlnProThrAsnProPheLeuGluMet-----IleLysPheLeuLe 605
      :
      :
      :
Db      2039 GATGAAGAAAGAAATTAAGAGACTGCGATGAGAGTGTCTTAACACAAATTAAGAGCTCA 2098
      :
      :
      :
Qy      605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
      :
      :
      :
Db      2099 TGAGAGAGTGGAGCCCTGCATTAATGACCAAGAACCTGTAAAGCC-----AAAGAGCA 2152
      :
      :
      :
Qy      625 lAsnLysSerIleAspGlyThrAlaAspAspGluValProThrAspGlnAl 645
      :
      :
      :
Db      2153 GAATCTTGTGT-----AGTCAAGTAAAGTGTCTTGAAGCTTGAAGAGGC 2194
      :
      :
      :
Qy      645 a---lLeArgAlaGluLeuGluLeuLeuValLeuSerPheThrHisProIleSerPh 664
      :
      :
      :
Db      2195 TCAGTTCGTACAGAGCCCTTGATGAGGCCAA-----AATATTTATTTGTTT 2242
      :
      :
      :
Qy      664 eHisSerAla-----GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
      :
      :
      :
Db      2243 GCAATCTTCAGTGAATGCGCTCATTCAGAGAGTAGAAGATGAGCAGAAACTGAGAGA 2302
      :
      :
      :
Qy      679 tAspAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysI1 699
      :
      :
      :
Db      2303 GAAGGATGAGAAACATCAGTACAGTGAATAATCAATTT----- 2339
      :
      :
      :
Qy      699 eGluIuLysPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis-- 716
      :
      :
      :
Db      2340 -----CAAGACCAAGAGCAGCTTGTCTTAAGTGTCCAGGTGGAAGAGACACCAACT 2395
      :
      :
      :
Qy      717 ---LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
      :
      :
      :
Db      2396 TTGGAAGAGCAAAACCTTAGAAACTGAGAAATCTGACATGGAATGAGAGCAAGAGATCA 2455
      :
      :
      :
Qy      735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
      :
      :
      :
Db      2456 AGTGCTACAAATCCAAAAATGCGCTTTCAGAGACACATTAAGAGTGTGCGAGAGTCTTA 2515
      :
      :
      :
Qy      755 uAspProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLe 775
      :
      :
      :
Db      2516 C-----AAGATCTAGAGAT-----GAGCTTGATTT 2542
      :
      :
      :
Qy      775 uLeuAlaProAspGlnPheAlaAlaProTPrLysSerTyrValAlaThrPheIleValLy 795
      :
      :
      :
Db      2543 GACAAAAATGACAAAAATG-----TCCTTTGT----- 2570
      :
      :
      :
Qy      795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTyrValProAs 815
      :
      :
      :
Db      2571 -----GAAAAAGTAAACAAAAAGACTGCAAAAGCA 2599
      :
      :
      :
Qy      815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyr 835
      :
      :
      :
Db      2600 AACTGAGCTGCAGAGAGAAATGACATGAGATGCGACAGAAACAGCAGAGCTGCAGAGAGA 2659
      :
      :
      :
Qy      835 pLeuLeuGlyMetLysAsn-----AsnHisSerLy 845
      :
      :
      :
Db      2660 ACTCAGTGAAGAAATTAAGCTGAGTGCAGAGTTCGAGTACTGTGGAAGAAATGAAA 2719
      :
      :
      :
Qy      845 sSerGlyThrSerThrLeuAlaGlyLeuThr----- 855
      :
      :
      :
Db      2720 GAGCAGCAAAAGATCAATTAAGAGAGCTCACTAGCAAAATAGTGAATTAAGAGAGAGCT 2779
      :
      :
      :
Qy      856 -ThrIleuHisSerAspGlyAspLeuThrGluGlnIlyLysIleSerLysProAspMe 875
      :
      :
      :
Db      2780 AGATTGCATGCAGCAAGAC-----CAGGTGAAAGAGAGAGGAAAGT-----AAGAGAGAAAT 2833
      :
      :
      :
Qy      875 tSerArgLeuArgLeuAlaIleArgIleSerAlaIleValLysLeuAlaGlnIuProCysTy 895
      :
      :
      :
Db      2834 AGCTGAATATCAGCTACGCGCTTCATGAGAGCTGAAAGAAA----- 2873
      :
      :
      :

```

OY	895	RHISGILLETLEFTRLEUGLINTYGLILEUCYSALALEUALALEASAPGLUC	915
Db	2874	-CACACAGCTTGGCTTTTGGAC-----ACAAACAAACA	2905
OY	915	STYGTGLNALARG--GLNALPHEALGLNLSLEU-----HISLYSGILYLEUSE	931
Db	2906	GTATGAAGTAAGAAATCCAGACATACCGAGAGAATTGACTTCTTAAGAAGAATGCTCCAG	2965
OY	931	ARGLEUALARGLEUPROLEUGLUTYRMETALILECYS-----ALALE	945
Db	2966	TYTCACAGAGAGCTGGAGATAGACTTTTAAAGTCTAGTAAGAAGAGCTCAATAATTCATT	3025
OY	945	WCYSALALAYSAPPROVALYLSGULARGARGALAHISALARGINCYSLEUALAYSAS	965
Db	3026	GAAGACTTCTACTCAGATTGTTGGAGAAATTGACAGAAAACCAAGATGACAACTTAAATA	3085
OY	965	NILEASNALARGARGLUTYRLEULYSGLINHISALALAYALSERGILYLSLEULUSE	985
Db	3086	TGTAAT-----CAGTTGAGAAGAGAAATGACGTGCCAGGGGAAATGTAA	3133
OY	985	RLEULENPROGLUTYRVALVALPROTYRTHILEHISLEULenalAHISAPPROASPTY	1005
Db	3134	GTGTGTG-----	3144
OY	1005	RVALLYVALGLNASPILEGUINLEULYSAPVALYLSGLUCYSLEUTRPHEVALLE	1025
Db	3141	-----ATCAATCTCTTAACACACTGGAAGAGAAAAGAGATCTGCAGAAAAGAACT	3199
OY	1025	UGLULEULenalALAYSASNGLUASNASNSERHISALAPHEILARG-----	1041
Db	3194	CTCTCAACTTCAAGCTGCACAGAGAACAAACAAACAGTACTGTTAGTATACCAAGT	3253
OY	1042	-----LYSMETVALGLUASNILEYSGINTHLYSAPPLAINGLYPROASAPALALY	1066
Db	3254	CGATGAATTAACAACTGAGATCAAGAAGCTGAAGAAACCTTGAAGAAAAACCAAGGA	3313
OY	1060	MEETASNLULYSLEUTYRTHVALCYASAPVALALAMETASNILEILEMETSERLYSE	1080
Db	3314	GGCAGATCAATACTTGGATAGTACTGT-----TCCTTGCTTATAGGCATGA	3361
OY	1080	RTHTHTRYISERLEUGLUSERPROLYSASP-----PROVALLENPROALARGPHEP	1098
Db	3362	A-----AAGTTAGAGAAAGCTAAAGAAATGTTAGAGACACAAAGGCCCATCTGTG	3412
OY	1098	ETHGGLNPROASPLYSASNPHESERASTHTRYLYSASNYRLEUPROPROGLUMETLYSE	1118
Db	3413	TTTCACAGCAATCTTAACAAGATTCGCCAGGGTCTCCTTGTGATGCTCAAGTTGT-----	3467
OY	1118	RPHEPETHRPROGLYLSPROLYSTHTRYASNVALLEUGLYALVALASNLYSPROLE	1138
Db	3468	-----CCAGAGCAACTCTCCAACTCCCTCTGTACTGAAAAGAGTT	3508
OY	1138	USERSERIALGLYLSGLINSERGLNTHTRYLSERSERARGMETGLUTHVALSERASNAI	1158
Db	3509	ATCTATCTTGCCCAAAATAAAGCTTCAGGCCAAAGGCCAAAGCTGGAATAATGGGAAGA	3568
OY	1158	ASERSESERSERASNPROSESERPROGLYARGILEYSGLYARGLEUASPSERSEGL	1178
Db	3569	TGTGTGGAGGACCAACACTGCTTACCCAGAGACCTTTTCTAAAAAAGCAAGAAAGCAGT	3628
OY	1178	UMETASP-----HISSEGLUASNGLUASPTYRTHMETSESERPROLEUPROGL	1199
Db	3629	CATGAGTGCTATTCACCTTCGACGAAGACAGGAAGGTACTAGTTTGAAGCCAGAGGACT	3688
OY	1195	YLYLSYLSERASPLYSARGASPSASPSERASPLEUVALARGSERGLULEUGLULYSPROAR	1215
Db	3689	TCCAGAAATTTTAAAGAAAGGTTTGTCTGACATCCCGACAGAAAGAACTACCCATATAT	3748
OY	1215	GGLYARGYLSYSTRPROVALTHNGLUGLNU-----GLULYSLEUGI	1230
Db	3749	CTGTGGAGGAACAACCATGCACTGGGAGCAGGCGCGCTGCTGTCACACAAGATTACC	3808

```

QY      1230 ymelaspspleuThr-----LysLeuValGlnIugLynsPro---LysGI 1245
      ::: |||||
Db      3809 CCTATCCCACTGACTGTCCCAACAAACAATCTTCGACAGATCTCCAAACCAACAGCTGG 3868
QY      1245 ySerGlnArgSerArgLys---ArgGlyHisThrAlaSerGluSerAspGlnGlnItr 1264
      |||||
Db      3869 TGGCAGCAGATCACCAAAAGTGTAAGTTGCTCAGCGCGGACCCAGTACTGAGCACCAT 3928
QY      1264 pProGlnIugLysArgLeuLysGlnAspIleLeuGlnAsnGlnLyspGlnGlnAsn----- 1282
Db      3929 CCTCCGAACACCCACACACGAATCCGTCCTCCACATATATCTTCCTGAGAGAAAGTCCGAC 3988
QY      1283 -SerProPolysLysGI-----LysArgGlyArgProProPolysProLeuGI 1298
      |||||
Db      3989 TGACAGACCCCAAGAGAGGGCCCTGAGGGCTCAACGCCCGCGCACTTGTCCCAACCCCAAGC 4048
QY      1298 yGly-GlyThrProLysGlnIugLupPothrMetLysThrSerLys 1312
      ||| ::|||
Db      4049 TGGACTGGAGTCCCAAGGCGACTGAGAACTGTAAAGGTCACAGTCAA 4092

RESULT 12
US-08-836-022A-10/c
? Sequence 102, Application US/08836022A
? Patent No. 6001357
? GENERAL INFORMATION:
? APPLICANT: Trustees of the University of Pennsylvania
? APPLICANT: Wilson, James M.
? APPLICANT: Fisher, Krishna J.
? APPLICANT: Chen, Shu-Jen
? APPLICANT: Weltzman, Matthew
? TITLE OF INVENTION: Improved Adenovirus Virus and
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESS: Howson and Howson
? STREET: Spring House Corporate Cntr, P O Box 457
? CITY: Spring House
? STATE: Pennsylvania
? COUNTRY: USA
? ZIP: 19477
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/836, 022A
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/331,381
? FILING DATE: 28-OCT-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Bak, Mary E.
? REGISTRATION NUMBER: 31,215
? REFERENCE/DOCKET NUMBER: GWPVN, 008PCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-540-9200
? TELEFAX: 215-540-5818
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 19307 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: CDNA
US-08-836-022A-10

Alignment Scores:
Pred. No.: 8,87e-06 Length: 19307
Score: 184.50 Matches: 241
Percent Similarity: 33.38% Conservative: 213
Best Local Similarity: 17.72% Mismatches: 496
Query Match: 2.56% Indels: 411

```

```

DB:      3          Gaps:      50
US-09-512-581b-2 (1-1391) x US-08-836-022A-10 (1-19307)

OY      206 AlaHisLysAsnLeuAsnLysGlnAlaIleuAspLeuAlaLysAlaLeuLeuLysArgThr 225
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      1204 GCTTCAGAAACACGTACAGACGGGTGGACAGAAATTCGTCCCAATTCGTGAGTGAAGAGT 11990
OY      226 Ala-----GlnAlaIleuGluProTyrIleThrIlePhePheAsnGlnValLeuMetLeuGly 244
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11989 AACTGGCTAGATGATCAACCAACATCATCTACCTTTATATACAGTCAACCAATTTGGAA 11930
OY      245 -----LysThrSerIleSerAspLeuSerGlu--- 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11929 CAGATGACACTACTGCCGCAAAACTTTGMAAAACCCAGTACCCCTATCAGAGCCA 11870
OY      254 HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSerVal 273
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11869 ACAGCAATTAAGCCAGTTAAATAATTTGTAAGATGAGTCAACAGATTGTACAGCTCTT 11810
OY      274 LeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVal 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11809 CAGCTCAAAATTTAG-----CAATTAATAATTCAGACTCA 11774
OY      294 LysLeuLeuAlaLysMetPheGlyValLysAspSerGluLeuAlaSerGlnAsnLysPro 313
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11773 CAACCTGAAAGAAAGGACAGGCG-----CCA 11747
OY      314 LeuTyrGlnCysTyrLeuGluYArgPheAsnAspIleHisValProIleArgLeuGluCys 333
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11746 ATGTTT-----CTGATGACAGCTTT 11726
OY      334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11725 GTGGCCTTACTAATCAT-----TTTAACCA 11699
OY      354 TyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 373
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11698 -----ATCTTTGAT 11690
OY      374 IleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
      ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11689 GGTGTAGGCGCCAAAGAGAAAGAGTACAGACAAATTTTGGACACTTACCACCAATGCCG 11630
OY      394 ArgGluArgThrLeuAspLysArgTrrArgValArgLysGluAlaMetMetLysLeuAla 413
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11629 TATCGAGAGCAATAGTAGC-----ATCAG 11603
OY      414 GlnIleTyrLysTyrAlaLeuGlnSerAlaIleAlaGlyLysAspAlaIleLysGlnIle 433
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11603 ----- 11603
OY      434 AlaTrrPheLys-----AspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAsp 451
      ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11602 ACGTGCATCCACAGTACAGAAAGCAAACTCTGTACCTTATCTAGTCTAGTAATAT 11543
OY      452 ArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThr 471
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11542 GAAATAATATGAGAGAGACTCGGAAATTTACAGGCTCCAAAGTTCTTTGAAA----- 11489
OY      472 GluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAla 491
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11488 GAGCAACAATAATGGCTTCAACTATCTGAGTACACTGTGAAGAGATGGCCAGAAAGCA 11429
OY      492 LeuAsnGluMet-----TrrLys 497
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11428 CTTTCAGAAATATGCCAGAAATATCTGCAGAAATTTGAAGAGATTTGAGGGCAGCTGGAAG 11369
OY      498 -----CysGlnAsnLeuLeuArgHisGlnValLysAsp 508
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11368 AAATCTTCTCCCACTGGTGGAAAGAGCTGCAAAAGCTAGAGAACAT----- 11321
OY      509 LeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

Db 10336 ATTGAAAGTCCCTTCACATTAAATTCAGAGAGTCCGTTGAATTCATTGACCAAGCACTGGCA 10377
 QY 820 ProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMet 839
 Db 10276 GCTTATTCACCTGACAGAGTGGATGCAGCTCAATGGCTTCAGAA-----GCC 10229
 QY 840 LysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThrIleLeuHis 859
 Db 10228 CAGAAATCCCAATCGAATTTTGACAACTGACAACTGAACTTATGAAAGAAATGCAAAACAT 10169
 QY 860 SerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMetSerArgLeuArg 879
 Db 10168 AACGAGGGAAGGATGCCAAACCAAGGCTTCTTCAACAATGTGATGTTCGACACAGAAAAA 10109
 QY 880 LeuAlaIleLysSerAlaIleValLysLeuAlaGlnGlnLysProCysTrpHisGln----- 897
 Db 10108 TTACAGATGCTCCACATGAATTTGCATTTATCCAAAACACAGCAATTTTGAACAAAGCT 10049
 QY 898 -----IleIleThrLeuGlnGlnTrpGlnLeuCysAlaLeu----- 909
 Db 10048 CTAGAGAAAGTAAAGATGATTTTATGATGAAGTCAAGTGCATTTGCGCTGCAATTCGAAACC 9989
 QY 910 ---AlaIleAsnAspGlyCysTrpGlnValArgGlnValArgLeuGlnLysLeuHisLys 928
 Db 9988 AAGAGTGTGAACAGAGAAATGATTCAGTCAACAAGTCACTGATGTGTGCACTGTATATAA 9929
 QY 929 GlyLeuSerArgLeuArg----- 934
 Db 9928 AGCCTGAGTGAAGTCAAGTCTGAAGTGAATGGTATTAACCCGAGCTCAAAATTTGTA 9869
 QY 935 -----LeuProLeu 937
 Db 9868 CAGAAAAAGCAGACAGAAAAATCCCAAGAGCTTGATGAAAGCAAGTGAACAGCTTTGCAAAATG 9809
 QY 938 GluTrpMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
 Db 9808 CATTAACATGAGTGGGTGGC-----AAGGTAACAGAGAAAGCAACAG 9764
 QY 958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlnTrpTrpLeuLysGlnHisAla 977
 Db 9763 TTGGAGAAATGCTTGAGATGTCGCCGTAAAGATGAGAAAGAA----- 9722
 QY 978 AlaValSerGluLysLeuSerLeuLeuProGluTrpValValProTrpThrIleHis 997
 Db 9721 -----ATGAATGCTCTTAACAGAAATGG----- 9701
 QY 998 LeuLeuAlaHisAspProAspTrpTrpValLys-ValGlnAspIleGlnGlnLeuLysAspVal 1017
 Db 9700 CTGGCAGCACAAGATACAGAAATGACAGAAAGATCAGACAGTGTG----- 9658
 QY 1017 LysGlnCys-----LeuTrpPheValLeu-----GluIleLeuMetAlaLysAs 1032
 Db 9657 -AAGGAATGCCAAGTAAATTTTGATTTCTGAAGTTGCCCTGGGGAAGAGCTACTCAAAAGAG 9599
 QY 1032 ngLAsnAsnSerHisIlePheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052
 Db 9598 ATTGAGAAACAG-AAGGCTCACTGGAAGAGCTGTACGAATTTAGGAAGTCTTGAAAT 9540
 QY 1052 palGlnGlyProAspAspAlaLysMetAsnGlnLysLeuTrpThrVal----- 1068
 Db 9539 GGTGTGGCGCAGAAAGAAACCTTGTAAGAAATAACTGAGTCTTGTGAACAGTAACTG 9480
 QY 1069 -----CysAspValAlaMetAsnIleIleMetSer----- 1078
 Db 9479 GATAGCTGTCACTCCAGAGTAGAAGAAATGCTAAATCTTTTGTGGATATACAGAAACA 9420
 QY 1079 ----LysSerThrThrTrpSerLeuGlnUserProLysAspProValLeuProAla---Ar 1096
 Db 9419 CATGGAACCTTTGATCAGAAATAGAAACAAATCACAAGTGGATCTATTATGACAGATGA 9360
 QY 1096 gPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTrpLeuProGluMet 1116
 Db 9359 ACITTTAGATGAGTCTGAAAAAGAAACCAACAAACAAAGAAAGACATTTCTTAAGCGTTT 9300

QY 1116 LysSerPhePheThrProGluLysProLysThrThrAsnValLeuGlnValAlaValAsnLys 1136
 Db 9299 AAGGCTGAAGTAATGATACATGCGCCCAAGGTGAGATCCACAGCTGACCAAGCAGAAAA 9240
 QY 1136 sProLeuSerSerAlaGly-----LysGlnSerGlnThrLys 1148
 Db 9239 ATTGATGGCAAAACCCGCGTGCACACCTGACAGAAAGTAGTAGAGCCCAAAATCTTGAGCT 9180
 QY 1148 sSerSerArgMetClnThrValSerAsnAlaSerSerSerSerAsnProSerProGln 1168
 Db 9179 CAACCGTGAATTTGACAGCTTATTTCTCACAGAAATTAAGACGTGAAAGGCTTCATTCCT-- 9122
 QY 1168 YArgGlyLeuGlyArgLeuAspSerSerGluMetAspHisSerGlnAsnGlnLysTrpThr 1188
 Db 9121 -----TGAAGGAATTTGAGACAGATTTAACTACATATATACAAAA 9084
 QY 1188 rMetSerSerProLeuProGluLys----- 1196
 Db 9083 ATTGCTTGAACCACTGAGGCTGAAATTCACAGCAGCGGCTGAATTCGAAAGAGAAAGCTT 9024
 QY 1197 -LysSerAspLysArgAspAspSerAspLeuValArgSerGlnLeuGlnLysProArgGln 1216
 Db 9023 CAATAAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8964
 QY 1216 YArgLysLysThrProValThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1236
 Db 8963 CAACCTTCAACAAAGAAATGATCAGATGAGAGAAAGCGAGAAATATGAATTAACACAGCA 8904
 QY 1236 sLeuValGlnGlnGln-----LysProLysGlySerGlnArgSerArgLysArg 1252
 Db 8903 GCTGTACAGCAAAAGATTAATGCTCAAGAGATTTGAGGCTCTCAAGAAAGAAAAAG-- 8846
 QY 1252 gGlnHisThrAlaSerGlnSerAspGlnGlnGlnTrpProGlnGlnGlnLysArgLeuLysGln 1272
 Db 8845 -----GCCCTAAGAAATTTCTCAACCTGATACAGTACAGTACAGTACAGTACAGTACAGT 8799
 QY 1272 uAspIleLeuGlnAsnGlnAspGlnGlnAsnSerProProLysGlnLysArgGlyArg 1292
 Db 8798 TGATCTCTGAAATGCTTGGATGAATTT----- 8771
 QY 1292 gProProLysProLeuGlnGlyGlyThrProLysGlnGlnProThrMetLysThrSerLys 1312
 Db 8771 ----- 8771
 QY 1312 sLysGlySerLysLysSerGlyProProAlaProGlnGlnGln-----GlnGln 1329
 Db 8770 -----GAAAAAAATTTACCAAGCCCTACCTGAACCCAGAGATGAAGAAATTTAAAGCA 8718
 QY 1329 uGlnGlnArgGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSer 1349
 Db 8717 AATTGATCGTGAATTTG-----CAGAAAGAAAGAAAGAGAGCTGAATCCAGTGCAG 8670
 QY 1349 rArgArgAlaGlnGlnArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1368
 Db 8669 CAGGCAAGCTGAGGCTTGTCTGAGATGGGCGCGCAGATGCGAGTGGAGCCAACTCAG 8612

RESULT 13
 US-09-427-048A-10/c
 ; Sequence 10. Application US/09427048A
 ; Patent No. 6203975
 ; GENERAL INFORMATION:
 ; APPLICANT: Trustees of the University of Pennsylvania
 ; Wilson, James M.
 ; Fisher, Krishna J.
 ; Chen, Shu-Jen
 ; Weltzman, Matthew
 ; TITLE OF INVENTION: Improved Adenovirus Virus and
 ; Methods of Use Thereof
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Howson and Howson
 ; STREET: Spring House Corporate Cntr, P O Box 457

CITY: Spring House
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19477

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/427,048A
 FILING DATE: 21-Oct-1999
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/836,022
 FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: Bak, Mary E.
 REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: GNP/PN.008PCT

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19307 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-427-048A-10

Alignment Scores:

Pred. No.: 8.87e-06 Length: 19307
 Score: 184.50 Matches: 241
 Percent Similarity: 33.38% Conservative: 213
 Best Local Similarity: 17.72% Mismatches: 496
 Query Match: 2.56% Indels: 411
 DB: 3 Gaps: 50

US-09-512-581B-2 (1-1391) x US-09-427-048A-10 (1-19307)

QY 206 AAlahlsysasnleuasnlysglnalatrarspleuallalyalaleuuleuylsargthr 225
 DB 12049 GCTTCAGACACGACGACGCCGCGGACAGCAATTCGCCAATTCGCTGAGAGAGCTT 11990
 QY 226 Ala--glnalailegluprotlyrlthrhrpheasnnglnvalleuuleuileugly 244
 DB 11989 AACGGGCTAGAGTCAACACACATCATTCCTTTTATATCAGCTCAACATTTGCA 11930
 QY 245 -----lysthrSerileSerAspleuSerlu--- 253
 DB 11929 CAGATGACACTACTGCCGAATACTGTGTGAACCCAGCTACACCCCTATCAGAGCCA 11870
 QY 254 HisValpheaspleuileuugluLeuylrnsnleaspserrHisleuLeuLeuSerVal 273
 DB 11869 ACAGCAATTTAAAGCCAGTTAAATTTGTAGATGAAGTCAACAGATTTGTCAGCTCTT 11810
 QY 274 leuproglnleugluPhelyleuLysSerasnspasnnglnugluargleuglnvalVal 293
 DB 11809 CAGCTCTCAATTGAG-----CAATTAAATAATTCAGAGTCTTA 11774
 QY 294 LysleuLeuallalyMetPheGlyAlalalysaspsergluLeuallaserGlnasnlyPro 313
 DB 11773 CAACCTGAAGAAAGGACAGGAG-----CCA 11747
 QY 314 leutrrpglnCysTyrleuglyArgPheasnspIleHisValproIleargleuglnCys 333
 DB 11746 ATGTTT-----CTGGATGACAGACTTT 11726
 QY 334 ValLysPhealaserHisCysLeuMetasnHisProaspleuallalyAspleuthrCln 353
 DB 11726

DB 11725 GTGGCTTTACTAATCAT-----TTTAACCAAC----- 11699
 QY 354 TyrleuLysValArgSerHisasProgluallalearghIsasPvalIleValSer 373
 DB 11698 -----ATCTTTCAT 11690
 QY 374 IleValThrAlaAlalalysAspIleleuLeuValasnAspHisleuLeuasnPheVal 393
 DB 11689 GGTGTAGGCGCAAGAGAGAGCTACAGACAAATTTTTCACCTTATCCCAACATGGCGC 11630
 QY 394 ArggluArgThrleuAspLysArgTrpArgValArgLysGlnalameleGlyLeuAla 413
 DB 11629 TATCAGAGACATAGTAGC-----ATCAGC----- 11603
 QY 414 GlnIleTyrLysTyrAlalaleuGlnSerAlalalaglyLysasPalaAlalysGlnIle 433
 DB 11603 ----- 11603
 QY 434 AlatrPileLys-----AspLysleuLeuHisIleTyrTyrGlnasnSerIleasPasp 451
 DB 11602 ACGTGATCCAGACAGTCAGAAAGCAAACTCTGTACTTATCTTAGTTCAGTGAATAT 11543
 QY 452 ArgleuLeuValgluArgIlePhealaglnTyrMetValProHisasnleugluThrThr 471
 DB 11542 GAATTAATGAGAGAGAGACTCGGGAATTTACAGGCTCGCAAGTCTTTGAAA----- 11489
 QY 472 GluArgMetLysCysLeuTyrTyrleuTyrAlarThrleuAspLeuasnAlalalysAla 491
 DB 11488 GAGCAACAATAATGGCTTCACTATCTGAGTACACTGTGAAGAGAGATGCCAAGAAACA 11429
 QY 492 LeuasnGluMet-----TyrLys 497
 DB 11428 CCTTCAGAAATATGCCAGAAATATCTGTCAAGATTTGAAGAGATTCAGGGCCACTGGAG 11369
 QY 498 -----CysGlnasnleuLeuArgHisGlnValLysasp 508
 DB 11368 AAATCTTCTCCCACTTGTTGGTGAAGCTGCCAAAAGCTGAGAAGACAT----- 11321
 QY 509 LeuLeuAspLeuileLysGlnProLysThrAspAlaserValLysAlaIlePheSerLys 528
 DB 11320 -----ATGAATTAACCTTCGAAAATTTTCAGATTCACATTAACAACTTA---CAGAAA 11273
 QY 529 ValMetValIleThrArgAsnleuProasProgluLysAlalaglnAspPheMetLysLys 548
 DB 11272 TGGAG-----GCTGAAGTTGATGTTTCTCGTGAAGAG 11240
 QY 549 PheThrGlnValleuGluAspAspGluLysIleargLysGlnleuGluValleuValSer 568
 DB 11239 GAATGGCCTCGCCCTGGGGATGCTGAATTCCTGAAAAACAGCTCAAA----- 11192
 QY 569 ProthrCysSerLysGlnAlalagluGlyCysValArgGluIleThrLysLysleugly 588
 DB 11191 -----CAATGCAGACTTTAGTTGTATATTCAAACAAATTCACCCAGTTTAAT 11141
 QY 589 AsnProLysGlnProThrAsnProPheleuLeuMetIleLysPheleuLeuGluArgIle 608
 DB 11140 AGT-----GTTAAAGAAGCTGGCGGAGAGATTAAGAGTGAAGCTGACTTTAG 11093
 QY 609 AlaProValHisIleAspThrGluSerIleSerAlaleuIleLysGlnValasnLysSer 628
 DB 11092 TTTCGATCCACACAGCAGACAGAA-----CTTAGAGACTTTACACCTAG 11048
 QY 629 IleaspglyThrAlasPaspGluaspGluGlyValProThrAspGlnAlaIleargAla 648
 DB 11047 TGGGATCACAATATGC-----CGCAGGTCTACACAGAAAGAGCCCTTAAAGCA 10997
 QY 649 Glnleuglu-----LeuLeuYsValleuSerPheThrHisProIleSerPhe 664
 DB 10996 GGTTCGATTAACACCCGTAAGCTTCACAAAAGATCTATCAGAGATGACATGATGATGACA 10937
 QY 665 HisSerAlagluThrPheGluSerleuLeuAlacysLeuLysMetasPaspGlu----- 682
 DB 10936 CAAGCTGAAGAAGATATCTAGAGAGATTTTGAATTAACCTCCAGATGAATTTACAG 10877

QY 683 -----LysValAlaGluAlaLeuGln----- 690
 Db 10876 ACTGCTGTGAAGAAATGAGAGACGTAAGAAGAGGCACTACAAAAGAACTAAAGTCG 10817
 QY 691 -----IlePheLysAsnThrGlySerLysIleGluGlnAspPheProHisIleArgSer 708
 Db 10816 AACTCCCTTACTAGACTGTGAATATGTAATAGCTCCAGCTCCACCCTCAGCACAAGAG 10757
 QY 709 AlaLeuLeuProValLeuHisHisLysSerLys----- 719
 Db 10756 GCCTTTAAAAAGAACTTGAAACTCTGCACCAACACTACCAATGGCTGTCCACCAGGCTG 10697
 QY 720 LysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSer 739
 Db 10696 AATGGAAATGCAAAACTTTGGAAAGAGTTGGGCACTGTGGCATGTGGCATGATTGTCTAT 10637
 QY 740 LysGlu----- 741
 Db 10636 TTAGAGAAACCAACAACTGGCTCATGAAAGTAATGAAACTTAAACCAATGGAAT 10577
 QY 742 -----ThrGlnPheAlaGlnIlePheGluProLeu----- 751
 Db 10576 GTTCCTGCAGAGCCTGAGGAATCAGTGAAGTGTGAATCTCTGAAATCTGATGCAT 10517
 QY 752 HisLysSerLeuAspProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGly 771
 Db 10516 CATTCAGAGAGCAACCAATTCAGATTCTGTATGCGACAGCATCTTACAGATGAGAGA 10457
 QY 772 HisIleAlaLeuLeuAlaProAspGln-----PheAlaIleProTyrLysSerTyr 788
 Db 10456 GTCATGATGATCACTGATCAATGAGAGCTTGAGACGTTGATTAATCTCTGTGGAGGAAC 10397
 QY 789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThr 808
 Db 10396 CATGAGAGGCTGTGAGAAACAAACTGCTTGAACAGATACCAAGTACCGCTGCCACAGAA 10337
 QY 809 ThrLys-----LeuTyrValProAspGluGlnValSer 819
 Db 10336 ATTGAAAGATCTCTGCACTTAATTCAGAGACTCGCTGTAATTCATTCAGACAGCTGTGCA 10277
 QY 820 ProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyrPheLeuGlyMet 839
 Db 10276 GCTTATATCACTGACACAGGTGATGAGCTCAATATGCTCAGCA-----GCC 10229
 QY 840 LysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHis 859
 Db 10228 CAGAAATCCAAATCAGATTGTGACACAGTCATGAGATATGTAAGAAATGAGAAACAT 10169
 QY 860 SerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArg 879
 Db 10168 AACCCAGGAGAGGATGCCCAACCAAGGGTCTTTCACAAATGTGATGTGCACAGAAAAA 10109
 QY 880 LeuAlaIleLysSerAlaIleValLysLeuAlaGlnLeuProCysTyrHisGlu----- 897
 Db 10108 TTACAAAGATGTCTCCATGAATTTCCGATTTCCAAAAACCAAGCAATTTTGAACAAAGCT 10049
 QY 898 -----IleIleThrLeuGluGlnIleTyrGlnLeuCysAlaLeu----- 909
 Db 10048 CTAGAGAAAGTAAGATGATTTTATGATGATGATCAAGTGCATTTTCCCTGCATTTGGAACC 9999
 QY 910 ----AlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLys 928
 Db 9988 AAGAGTGTGAACAGAAATTAATTCAGTACAACTAAGTCACTGTGTGAACCTTGTATAAA 9929
 QY 929 GlnLeuSerArgLeuArg----- 934
 Db 9928 AGCTGTAGTGAAGTCAGTCTGAAGTGAATGATGTATTAACCGGACGTCAAATTTGTA 9869
 QY 935 -----LeuProLeu 937
 Db 9868 CAGAAAAAGCAGACAGAAATCCCAAGAGCTTGATGACAGAGTAAACAGCTTTGAAATTTG 9809

QY 938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
 Db 9808 CATTAACATAGTGGGTGCG-----AAGCTAACAGAGAAAGCAACAG 9764
 QY 958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
 Db 9763 TTGAGAAATAGCTTGAAGTTGTCCGTAAGATGAGAAAGAA----- 9722
 QY 978 AlaValSerGluLysLeuSerLeuLeuProGluTyrValAlaProTyrThrIleHis 997
 Db 9721 -----ATGATGTCTTAAACAATGC----- 9701
 QY 998 LeuLeuAlaHisAspProAspTyrValLys-ValGlnAspIleGluGlnLeuLysAspVal 1017
 Db 9700 CTGCAGACAACAGATACAGAAATTAAGACAAAGATCAGCACTG----- 9658
 QY 1017 LysGluCys-----LeuTyrPheValLeu-----GluIleLeuMetAlaLysAs 1032
 Db 9657 -AAGAAATGCAAGTAATTTGGATTCTGAACTTGCTGGGGAAGGCTACTCAAAAGAG 9599
 QY 1032 nGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052
 Db 9598 ATTGAGAAACAG-AAGGCTCATCTGAAGATGTTACAGAAATGAGAGCTTTGAAAT 9540
 QY 1052 PAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrVal----- 1068
 Db 9539 GGTGTGGCGCAAGAAAGAAACCTTGGTGAAGATTAACCTGATCTTCAACAGTAAC 9480
 QY 1069 -----CysAspValAlaMetAsnIleIleMetSer----- 1078
 Db 9479 GATAGCTGTCACTCCAGAGTAGAAGAGTGGCTTAATCTTTGTGGAAATACCAAGACA 9420
 QY 1079 ----LysSerThrThrTyrSerLeuGlnIleSerProLysAspProValLeuProAla--Ar 1096
 Db 9419 CATGGAACCTTTGATCAGAAACATAGAACAAATCACAACAAAGTGAATCATATCACAAGTGA 9360
 QY 1096 gPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrIleuProGluMet 1116
 Db 9359 ACTTTTGAAGAGCTGCAAGAAAGAAACCAACCAACAAAGACATCTTAAAGCTTT 9300
 QY 1116 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuLysAlaValAsnLys 1136
 Db 9299 AAGGCTGAATGATATGATATGCGCCCAAGGTGAGACTCCACACGTGACCAAGCAGCAAA 9240
 QY 1136 sProLeuSerSerValGly-----LysGlnSerGlnThrLys 1148
 Db 9239 ATTGATGCAAAACCCGCGTACCACTGCAGAAAGTAGTAGGCCCAAAATCTGTAGCT 9180
 QY 1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGln 1168
 Db 9179 CAACCGTCGATTTTGCACGCTATTTCTCACAGAAATTAAGACAGGAAAGGCCCTCAATTCCT-- 9122
 QY 1168 YArgIleLysGlyArgLeuAspSerSerGluMetAsnHisSerGluAsnGlnAspTyrThr 1188
 Db 9121 -----TTGAAGAAATTTGAGACGCTTTAACTCAAGATATACAAA 9084
 QY 1188 rMetSerSerProLeuProGlyLys----- 1196
 Db 9083 ATTGCTTGAACCACTGAGGCTGAATTTACAGCAGGGGCTGAATCTGAAAGAGAGAGCTT 9024
 QY 1197 -LysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArgGln 1216
 Db 9023 CAATTAAGAAATATGATGAAACAAATAGAGGTACTGTAAATGATTTGTTGCAAAAGAGAGA 8964
 QY 1216 YArgLysLysThrProValThrGluGlnGlnGluLysLeuGlnLysMetAspAspLeuThrLys 1236
 Db 8963 CAACCTTACAACAAAGAAATCACAGATGAGAGAAAGCCGAGAGCAATTAAGATTAACAGCA 8904
 QY 1236 sLeuValGlnGluGln-----LysProLysGlySerGlnArgSerArgLysAsp 1252
 Db 8903 GCTGTTACAGACAAACATATATGCTCTCAAGAGTATTAGAGCTTCAAAGAAAGAAAAAG-- 8846
 QY 1252 gGlyHisThrAlaSerGluSerAspGluGlnGlnIleTyrProGluGlnLysArgLeuLysGln 1272

```

Db      8845 -----GCCCCAGAAATTTCTCACAGTGGTATCAGTACAGAGCAGCGCTCA 8799
      :      :      :      :      :      :      :      :      :      :
Qy      1272 uasplleuenglunasngluaspgluInasnsrProProlysLysgLyLysargGlyAr 1292
      :      :      :      :      :      :      :      :      :      :
Db      8798 TGATCTCTGAAATCTTGATGAAATG----- 8771
      :      :      :      :      :      :      :      :      :      :
Qy      1292 gProProlysProleuGlyGlyGlyThrProLysgluInuProThrmetylThsery 1312
      :      :      :      :      :      :      :      :      :      :
Db      8771 ----- 8771
      :      :      :      :      :      :      :      :      :      :
Qy      1312 slsGlyserLysLysSerGlyProProlaProgluInu-----gluGl 1329
      :      :      :      :      :      :      :      :      :      :
Db      8770 -----GAAAAAAATTATGACCGACCTTACCTGACCCAGATGAAAGAAATTAAAGCA 8718
      :      :      :      :      :      :      :      :      :      :
Qy      1329 uclugluArgInserGlyAsnThrGluGluInLysSerLysGlnHisArgValse 1349
      :      :      :      :      :      :      :      :      :      :
Db      8717 AATTGATCGTGAATG-----CAGAAAGAAAGAGAGCTGATGACAGTGGC 8670
      :      :      :      :      :      :      :      :      :      :
Qy      1349 rArgArGlaInGlnArgAlaGluSerProgluInserAlaIleGluSerThrGln 1368
      :      :      :      :      :      :      :      :      :      :
Db      8669 CAGGCAAGCTGAGGCTTGCTGTGAGAAATGGGCGCCCAATGCGAGTGGACCACTGAG 8612
      :      :      :      :      :      :      :      :      :      :
RESULT 14
US-08-195-487-3
: Sequence 3, Application US/08195487
: Patent No. 5783403
:
: GENERAL INFORMATION:
: APPLICANT: TOURKATLY, GARY
: APPLICANT: LIDGARD, GRAHAM P
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: TESTA HURWITZ & THIBEAULT
: STREET: 53 STATE STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/195,487
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/901,701
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7100
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6306 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..6306
: PUBLICATION INFORMATION:
: AUTHORS: SAILAK, ILLYA
: AUTHORS: CLEVELAND, DON W

```

```

: TITLE: PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
: TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
: TITLE: SEGREGATION OF PROTEINS AT MITOSIS
: JOURNAL: JOURNAL OF CELL BIOLOGY
: VOLUME: 116
: PAGES: 1395-1408
: DATE: MAR-1992
US-08-195-487-3
Alignment Scores:
Pred. No.: 1,9e-06 Length: 6306
Score: 182.50 Matches: 281
Percent Similarity: 31.34% Conservative: 195
Best Local Similarity: 18.50% Mismatches: 549
Query Match: 2.54% Indels: 494
DB: 1 Gaps: 55
US-09-512-581b-2 (1-1391) x US-08-195-487-3 (1-6306)
Qy      16 ProProGlyValLysGluLieserAspLysLieserLysGluGluMetValArgArgLeu 35
      :      :      :      :      :      :      :      :      :      :
Db      823 CCCAAGAGAGCTTGAGAGAGCTGCGTGACAAG-----AATGAGAGCTTACATGCGGCTG 876
      :      :      :      :      :      :      :      :      :      :
Qy      36 LysMetValValLysThrPheMetAspMet----- 45
      :      :      :      :      :      :      :      :      :      :
Db      877 CATGAACCCCTGACAGACGTCAGACACCTGGAAGACAGAGACAGCAGATGATGCGAAA 936
      :      :      :      :      :      :      :      :      :      :
Qy      46 ---AspGlnAspSerGluGluGluLysGluLeuTyrlsAsnLeu-----AlaLeu 61
      :      :      :      :      :      :      :      :      :      :
Db      937 ATCAACACAGCTTTCGAGAGAGAAATGAGACCTTTCCTTAACCTGCGGAGTTGCACT 996
      :      :      :      :      :      :      :      :      :      :
Qy      62 HisLeuAlaSerAspPheLeuLysHisProGlyLysAspValArgLeuValAla 81
      :      :      :      :      :      :      :      :      :      :
Db      997 CATCTG----- 1002
      :      :      :      :      :      :      :      :      :      :
Qy      82 CysCysLeuAlaAspIlePheArgIleTyrlaProGluAlaProTyrlThsProAsp 101
      :      :      :      :      :      :      :      :      :      :
Db      1003 -----CAG 1005
      :      :      :      :      :      :      :      :      :      :
Qy      102 LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLys 121
      :      :      :      :      :      :      :      :      :      :
Db      1006 CAGCTACAGAGATGCCCTCAATGACTGACGAG-----GAGCAGACGACAG 1050
      :      :      :      :      :      :      :      :      :      :
Qy      122 SerProGln-----PheAsnArgTyrlThrLeuLeuGluAsnIleAlaTrpValLys 139
      :      :      :      :      :      :      :      :      :      :
Db      1051 GCCACTCAGAGCTGCTGAGAGACAGGCCCGCAGGAGAGAGACTGACGCGACCCCTG 1110
      :      :      :      :      :      :      :      :      :      :
Qy      140 SerTyrlAsnIleCysPheGluLeuGluAspSerAsnGluIle-----PheThr 155
      :      :      :      :      :      :      :      :      :      :
Db      1111 CAGGCAAGAAATGCT-----CTTGAAGAGAAAGAAATCTCTCAGGAAACTTTCA 1164
      :      :      :      :      :      :      :      :      :      :
Qy      156 GlnLeuTyrlArgThrLeuPheSerValIleAsnAsnGlnLysAsnGlnLysValHisMet 175
      :      :      :      :      :      :      :      :      :      :
Db      1165 CAGCTGGAAGAACACTTGCCAGCTGCGAGTATACACCCAGAGAGAGGCGAGG 1224
      :      :      :      :      :      :      :      :      :      :
Qy      176 HisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnIleLeu 195
      :      :      :      :      :      :      :      :      :      :
Db      1225 -----CTGGGAGATGCTTGCAGCTGGAAC 1251
      :      :      :      :      :      :      :      :      :      :
Qy      196 LeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrl 215
      :      :      :      :      :      :      :      :      :      :
Db      1252 TTGAAGCAAGAGGACGACACTTCTGCTGCAAAACACACAGACTCCMAAGCGGTACAG 1311
      :      :      :      :      :      :      :      :      :      :
Qy      216 Aspleu-----AlaLysAlaLeuLeuLysArg----- 224
      :      :      :      :      :      :      :      :      :      :
Db      1312 ATGCTGAGACTGACGAGCGACGAGCAAGAGAACTGCTTGAGAGGGGCGACACTTC 1371
      :      :      :      :      :      :      :      :      :      :
Qy      225 -----ThrAlaGlnAlaIleGluProTyrlIleThrThrPhePheAsnGlnValLeuMet 242
      :      :      :      :      :      :      :      :      :      :
Db      1372 GAAGAAAGAAAGCAGCAGCTGCTGAGCTGATGACTGACTG----- 1413
      :      :      :      :      :      :      :      :      :      :
Qy      243 LeuGlyLysThrSerIleSerAspLeuSerGlu----- 253
      :      :      :      :      :      :      :      :      :      :

```

Db 1414 -----CAGAGCTCCATCTCCAACTCAGCCAGGCCAAGAGAGCTGGAGCAGGCTCC 1467
 QY 254 -----HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu----- 269
 Db 1468 CAGGCTCATGGGGCCGGTTGACATCGCCAGGTGGCCCTCTGACCTCTGAGCTCAGACACA 1527
 QY 270 LeuLeuSerValLeuProGlnLeuGluPheLys-----LeuLysSerAsnAspAsn 286
 Db 1528 CTCAAATGCCACATCCAGCAAGAGATCAAGAACTGGCTGCTGAGAGCAGCAGGCCAAA 1587
 QY 287 GluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
 Db 1588 GAGAGAGAGGCCCGCAGTACGACAGACCCCTC-----CAACAGCAAGAA 1629
 QY 307 LeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGluArgPheAsnAspIleHis 326
 Db 1630 CAGGCTCCAGGCGC----- 1644
 QY 327 ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
 Db 1645 -----CTCCGCCACAGGTGGAGCAGCTAAGCAGTACGCTGAGAGCAGAAAGCAGCAG 1698
 QY 347 LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
 Db 1699 TTG---AAGAGGTAGCGGAG-----AAGCAGAGGCACT 1731
 QY 367 ArgHisAspValIleValSerIleValThr----- 376
 Db 1732 AGGAGAGACCATGCCCAAGCAACTGGCAGTCTGTCAGAGAGCAGAGGCGCTCCTTAAG 1791
 QY 377 ----- 380
 Db 1792 GAGCGGATCGCGCTCTCAAGCAGCTGAGCAGTGGAGAGAAAGGCGTCCCAAGCTG 1851
 QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluAlaGluThrLeuAspLys 400
 Db 1852 GAGATTGTGACAGCAGCACTCAGGTGCTATGATAGCGCGGACAGCTGCCACAGACTCA 1911
 QY 401 ArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr-- 419
 Db 1912 GTGACACAGGCCCGCGGAGAG-----GACAGCTGACGCCGGAAGGTGGAG 1959
 QY 420 AlaLeuGlnSerAlaAla-----GlyLysAspAlaAlaLysGlnIle 433
 Db 1960 GAATCTCCAGCGCTGTGTGAGACAGCCCGCAGAGAACAGCATGAGGCCAGCGCCAGGTT 2019
 QY 434 AlaTyrIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArgLeu 453
 Db 2020 GCAGAGCTAGAGTTGCACTGCGCTGTGAGCAGCAAAAGCAACTGCAAAAGAAAGGCTG 2079
 QY 454 LeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArg 473
 Db 2080 GCCCAGAGAGAAAGCAAGCTCCAGAGCAGCTCCAGGCCCTCAAAAGAGTCTTGAAGTTC 2139
 QY 474 MetLysCysLeuTyrTyrLeuTyrAlaThrLeuAsp-----LeuAsnAlaValLys 490
 Db 2140 ACCAAG-----GGCAGCCTTGAAGAGAGAGAGAAAGCCGAGCGCTGCAGAT 2181
 QY 491 AlaLeuAsnGluMetTyrLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 510
 Db 2182 GCCCTGGAGAGACGACGCGCTGTATCTGTGAGCTGAAAGCAGAGACCCGGAAGCTGGTG 2241
 QY 511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
 Db 2242 GAGCAGCATTAAGCGGAAGAAAGAG----- 2268
 QY 531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550
 Db 2268 ----- 2268
 QY 551 GluValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThr 570
 Db 2269 -----CTGAAAGAAAGAGAGGCGCTGGCGCAAGGCGCTGAGGCTCGATTACTCAGACTT 2322

QY 571 CysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGluAsnPro 590
 Db 2223 GGGAGAGCCCATCAGCGCTGAG-----ACTGAAGTCCCGCG----- 2358
 QY 591 LysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluAlaGluIleAlaPro 610
 Db 2359 -----CGGAGGCTGGCAGGAGGCAATGGCTGCC 2385
 QY 611 ValHisIleAspThrGluSerIleSerAlaLeuLeuGlnValAsnLysSerIleAsp 630
 Db 2386 CAGCACACAGCTAGAGAGTGAAGTGTGACAGCAGCTGTCTAAAGAGTAGTCCCTGGCGTAC 2445
 QY 631 GlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
 Db 2446 GGGTATGAGGATACGACCAAGAGAGAGCAGCAGATGAGCGGCATGTTCCAGAGACAGCTG 2505
 QY 651 GluLeuLeuLysValIleSerPheThrHisProIleSerPheHisSerAlaGluThrPhe 670
 Db 2506 ATGACTTTGAG-----GAGGAATGT 2526
 QY 671 GluSerIleLeuAlaCysLeuLysMetAspAspGluLysValAla-----Glu 686
 Db 2527 GAGAGGCCCGCCAGGAGCTGCAGAGCAGCAAGAGAGTGGCAGGCATGAATCCAC 2586
 QY 687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
 Db 2587 AGCAGAGCTCCAGATTAAGCCGCGACAG-----AACAAACTAGTGTAGTCTCAATGCCAACCTG 2643
 QY 707 ArgSerAlaLeuLeuProValLeuHisIleLysSerLysGlyProProAlaGluAla 726
 Db 2644 GCCAGAGCACTCCAGCAGCTC----- 2664
 QY 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
 Db 2665 -----CAAGAGAAAGAGATCAGGCGCCAGAAAG 2691
 QY 747 IlePheGluProLeuHis-----LysSerLeuAspProSer 758
 Db 2692 CTTCAGATGACCTCTCCACTGTGCAAGAAAGATGGCTGCACAGCAAGAGAGTGGCC 2751
 QY 759 AsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuAlaPro 778
 Db 2752 CGCTTGAGAGACTTGGTC----- 2769
 QY 779 AspGlnPheAlaAlaProTyrLysSerTyrValAlaThrPheIleValLysAsp--Leu 797
 Db 2770 -----CGCAAGCGCAGGTGACAGCAGAGAAACAGCTCCCGGAGTTAGTCAAGAGAGCTGCG 2826
 QY 798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTyrValProAspGluGlu 817
 Db 2827 AGGCGAGAGACAGACAGCCGAG-----TGCGTG-----GAAGAG 2862
 QY 818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyrPheLeu 837
 Db 2863 CAACAGGAGCGCCAGTTCTGCAGACACAGCGCAG-----CTTCAG 2904
 QY 838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIle 857
 Db 2905 GCTATGAGCGGAGGAGAGACAGACATGGCAATAGCTGCAAGCGCTGCGGCGCGCTG 2964
 QY 858 LeuHisSerAspGluAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
 Db 2965 ATGAGAGCGCAGGCGCAGCAGACAGAGAGAGCGTGGCAGCAGAAAGAGAGTGGCGCG 3024
 QY 878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
 Db 3025 CTGACCCAGAGCGGCGCGCTGCCAGCTGACCTGCCTCGAG----- 3069
 QY 898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917
 Db 3069 ----- 3069

OY	918	ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu	937
Db	3069	-----	3068
OY	938	GLuTYrMetAlaIleCysAlaLeuCysAlaSerProValLysGluArgAlaHis	957
Db	3070	-----	3070
OY	958	AlaArgGlnCysLeuValLysAsnIleAsnValArgArgLysTYrLeuLysGlnHisAla	977
Db	3088	CTTGAGATGGCGGTGCGACAAAGCCCTCAACAGCAGCGCT-----GTGGAGTTCCT	3138
OY	978	AlaValSerGluLysLeuLeuSerLeuLeuProGluTYrValValProTYrThrIleHis	997
Db	3139	ACCCGTGGAGAGGCGCTGCTCATGCGCTGACGGAAAGAA-----	3180
OY	998	LeuLeuAlaHisAspProAspTYrValLysValGlnAspIleGlu-----GlnLeuLys	1015
Db	3181	-----GGCAAGGACACGAGACTTGGCCAGAGCTGTGGTCTGGAGGCGACCCAGATAA	3234
OY	1016	AspValLysGluCysLeuTYrPheVal-----LeuGluIleLeuMetAlaLysAsnGlu	1033
Db	3235	GAGCTGGAGAACTTCGGCAAAACCGTGAAGCACTGAAGAAAGAAAGCTGCTAAGAAAGAA	3294
OY	1034	AsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAla	1053
Db	3295	AAGGAGCAGCGACTGGCTGCAGAGCCCAATCTGAGCGCTGCTGGACAGACAG-----	3348
OY	1054	GlnGlyProAspAspAlaLysMetAsnGluLysLeuTYrThrValCysAspValAlaMet	1073
Db	3349	-----CCACAGGCCCCACAGCTGAGACACTCGCGGCGACAGGTGAGCAAGCTGGAACAG	3402
OY	1074	AsnIleIleMetSerLysSerThrTYrTYrSerLeuGlnSerProLysAspProValLeu	1093
Db	3403	CAATGCCAGAGACAGACAGACAGCGCTGACAGCCTGGAAACCAAGCCTGAGAGCTGAGCGG	3462
OY	1094	ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTYrLeuPro	1113
Db	3463	GCCCTCCCGG-----GCTGAGCGGGGACAGTCTCTGAGACTCTGACAGGGCCCATGTAGAG	3516
OY	1114	ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAla	1133
Db	3517	GAGAAAGCCCCAGGAG-----CTAGGGGCAC	3540
OY	1134	ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGln	1153
Db	3541	AGTCAGAGTGGCTTAGCCTCGGCCCAACGGGACGTGGCTGCTCCGACCAAGAGTACAA	3600
OY	1153	uThrValSerAsnAlaSerSerSerAsnProSerSerProGly-----ArgIle	1170
Db	3601	GACCACAGCAAGCGCTGAAGATGATGTGGAAGCCACAGTGGCCCGGGCGGCAAGAGCT	3660
OY	1170	elyGlyArgLeuAspSerSerGluMetAsp-----HisSerGluAsnGlu	1186
Db	3661	GAGAGGAAATAATACCTCATCAGACGCTTGAGAGAGAGAGTGTCTCATCTCTAA-----	3713
OY	1186	sPTYrThrMetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAsp	1205
Db	3714	-----TCCGACAGTCTCTGAGAAAGAGAGGGGAGACCAAGCAAGACTTGACCGG	3759
OY	1206	LeuValArgSerGluLeuGlnLysProArgGly-----Arg	1217
Db	3760	CTGGTGAATGGCCGAGTCAGAGAAAGCCAGAAAGCTGAGAGAGACTGGCGCTGCTGCAGG	3819
OY	1218	LysLysThrProValThrGlnGlnGlnGluLysLeuGlyMetAspAspLeuThrLysLeu	1237
Db	3820	CAGAGACAGCCAGCAACAGTCCAGACTGCAG-----AACGCACTCTG	3864
OY	1238	ValGlnGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSer	1257
Db	3865	CTCTGCGGGAGAGAGTGCACAGCTCCGGGAGGAGAGCTGAGAAACACGGGTGGCTTCA	3924
OY	1258	GluSerAspGlnGlnGlnThrProGluGluLysArgLeuLysGluAspIleLeuGlnAsn	1277

```

Db      3925 GAGAACCTCGCGGACAGGACTACCTCACAGGCTGAGCCTGGCGAGGACTG-----GGC 3978
Qy      1278 GLuSPgluGlnAsnSerProProLys-LysGlyLysArgGlyLarProProLysProLe 12978
      :::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      3979 CAAGAATTGAAGGGCGTGGCAGGAGAAgTTCTTCAGAAAGACAGAGCCCTCTCCACCTG 4038
Qy      1297 uGlyGlyGlyThrProLysGluGluPro-----ThMetLysThrSerLysLys 1313
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      4039 CAGCTCGAGCACACACAGCACAGGCCCTGGTGAAGTGAAGTCACTGCTGCTGCTGCTGCTG 4098
Qy      1313 sGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGluGluGluGlu 1330
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      4099 TGCCAGCAGCTGCAGGCCGAGCAGGCGCCTCCAGAAACCCACCGTGAAGACTGGAG 4158
Qy      1330 ----- 1330
Db      4159 CAGAGCAAGCAGCGCGTGGGAGACTGCGGCGCAGAGCTGCTGCGGCCACAGGAGACTT 4218
Qy      1331 -----GluArgGlnSerGlyAsnThrGluGlnLysSerLys 1343
      4219 GGGAGACTGATTCTCTGCGGCAAGAGTGGCAGACAGCAGCCAGACAGCTCACACACTG 4278
Qy      1343 rLySGlnHisArgValSerArg-ArgAlaGlnGlnArgAlaGlnSerProGluSerSera 1363
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      4279 CGGGCAGAGGAAGCGCAGCTATGCAGAGAGAGCTGAGCATGCTGGAAGAAGCCCATGGCCTG 4338
Qy      1363 laIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      4339 CT-----GGCAGAGGAGAACCG 4355

RESULT 15
PCT-US93-06160-3
; Sequence 3, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7100
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:

```


QY 551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThr 570
 Db 2269 -----CTGGAAGAGAGAGGGCTGGCGCAAGGGCGAGCTCGATTACTGACGCT 2322
 QY 571 CysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro 590
 Db 2333 GGGGAGGCCCATCAGCTGAG-----ACTGAAGCTCTGGCG----- 2358
 QY 591 LysGlnProThrAsnProPheLeuGluMetIleLysPheLeuGluArgIleAlaPro 610
 Db 2359 -----CGGAGCTGGCAGAGAGGCCATGGCTGCC 2385
 QY 611 ValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
 Db 2386 CAGCACACAGCTGAGAGTGAAGTGAAGCAGCTCGCAAGAGTAGTGGCTGGGGGTGAC 2445
 QY 631 GlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
 Db 2446 GGGTATGAGATAGCCAGCAGAGAGGAGCAGATAGGCCCATGCTTCCAGACAGAGCTG 2505
 QY 651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPhe 670
 Db 2506 ATGACTTTGAC-----GAGCAATGT 2526
 QY 671 GluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAla-----Glu 686
 Db 2527 GAGAAAGCCCGCCAGAGAGCTGAGAGGCAAGAGAGAGTAGAGGCTAGATGCCAC 2586
 QY 687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
 Db 2587 AGCGAGCTCCAGATAGCCAGGCGCAGAG-----AACAACTAGCTGAGCTCCATGCCAACCTG 2643
 QY 707 ArgSerAlaLeuLeuProValLeuHisLysSerLysGlyProProArgGlnAla 726
 Db 2644 GCCAAGAGCACTCCACAGAGCTC----- 2664
 QY 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerLysGluThrGlnPheAlaGln 746
 Db 2665 -----CAAGAGAAAGATGACGGCCCGCAAG 2691
 QY 747 IlePheGluProLeuHis-----LysSerLeuAspProSer 758
 Db 2692 CTTCGACATGACCTTCACATCTGCAGAAAGATGGCTGCCACGCAAGAGAGTGGCC 2751
 QY 759 AsnLeuGlnHisIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPro 778
 Db 2752 CGCTTGGAGACTTGCTC----- 2769
 QY 779 AspGlnPheAlaAlaProThrLysSerTyrValAlaThrPheIleValLysAsp---Leu 797
 Db 2770 ---CGCAAGCGCAGTGAAGAGAGAAACAGCCCTCCGGAGTTAGTCAAGAGAGCTGGC 2826
 QY 798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTyrValProAspGluGlu 817
 Db 2827 AGGGCAGAGACAGACAGCCCGAG-----TGGCTG-----GAGAG 2862
 QY 818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyrPheLeu 837
 Db 2863 CAACAGGAGAGCCACTTTCAGCAGCACAGAGCAGG-----CTGCAG 2904
 QY 838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
 Db 2905 GCTATGAGCGGAGGAGCAGAGACAGATGGGCAATGAGCTGGAAGAGCTCGGGCGCGCTG 2964
 QY 858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
 Db 2965 ATGAGAGCCAGGGGAGCAGCAGAGAGAGCGCTGGGAGCAGAGAAAGAGAGTGGCGCG 3024
 QY 878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
 Db 3025 CTGACCCAGAGAGGGGGCGTGGCCAGGCTGACCTTGGCTGGAG----- 3069
 QY 898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917

Db 3069 ----- 3069
 QY 918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
 Db 3069 ----- 3069
 QY 938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
 Db 3070 -----AAGCGGCCAGACAG 3087
 QY 958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
 Db 3088 CTGAGATGGCGCTCAGAACGCCCTCAACAGCAGCAGCGT-----GTGAGTTGGCT 3138
 QY 978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
 Db 3139 ACCCTGCAGAGAGCACTGGCTCATGCCCTGACGGAAGAAAGAA----- 3180
 QY 998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGlu-----GlnLeuLys 1015
 Db 3181 -----GGCAAGACACAGAGAGTTGGCCAAAGCTTGGCTGTGGAGAGCCAGCCAGTATAAA 3234
 QY 1016 AspValLysGluCysLeuThrPheVal-----LeuGluIleLeuMetAlaLysAsnGlu 1033
 Db 3235 GAGCTGAGAGAACTTCGCCAAACCGTGAAGCACTGAAGAGAAACACAGCTGCTAAGAAAGAA 3294
 QY 1034 AsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAla 1053
 Db 3295 AAGGACAGCGATGTGGCTCAGAGCCCAATCTGAGGCTGCGCAGGACAGAG----- 3348
 QY 1054 GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValLysAspValAlaMet 1073
 Db 3349 -----CCAACAGGCCCAAGCTGGAACACACTCGCGCGAGAGTGAAGACAG 3402
 QY 1074 AsnIleIleMetSerLysSerThrTyrSerLeuGlnLeuSerProLysAspProValLeu 1093
 Db 3403 CAATGCCAAGACACAGAGAGCAGCTGACAGCTCGAAACGACGCTGAGCGCTGAGCGG 3462
 QY 1094 ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113
 Db 3463 GCCCTCCGG-----GCTGAGCGGAGCAGAGCTGCTGAGACTCTGACGGCCAGTTAGAG 3516
 QY 1114 ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuLysAla 1133
 Db 3517 GAGAAAGCCCAAGAG-----CTAAGGCAC 3540
 QY 1134 ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr-LysSerSerArgMetGlu 1153
 Db 3541 AGTCAGAGTGCCTTAGCTCGGCGCAACGGAGTTGGCTGCTCCGCCAACAAAGTACAA 3600
 QY 1153 ThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly-----ArgIle 1170
 Db 3601 GACCACAGCAGAGCTGAAGATGATGGAAGGCCAGTGGCCCGGGCGCGCAGAGAGCT 3660
 QY 1170 eLysGlyArgLeuAspSerSer-GluMetAsp-----HisSerGluAsnGluA 1186
 Db 3661 GAGAGGAAATAAGCTCATTCAGACGCTTGCAGAGAGAGGTCCATCTGAA----- 3713
 QY 1186 sPtyrThrMetSerSerProLeuProGly-LysLysSerAspLysArgAspSerAsp 1205
 Db 3714 -----TCGCCAGGTCTCGGAGAAAGAGGGGAGAGACAAAGAGTTGAAGCGG 3759
 QY 1206 LeuValArgSerGluLeuGlnLysProArgGly-----Arg 1217
 Db 3760 CTGGTATGATGGCGAATCAGAGAGCCAGAAAGCTGAGAGAGAGCTGGCGCTGACAGG 3819
 QY 1218 LysLysThrProValThrGluGlnGluLysLysLeuGlyMetAspAspLeuThrLysLeu 1237
 Db 3820 CAGACACAGCCAGCAACAGTGCAGAGCTGCAG-----AACGACCTCTG 3864
 QY 1238 ValGlnGlnLysProLysGlySerGlnArgLysArgGlyHisThrAlaSer 1257

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:21:35 ; Search time 769 Seconds

(without alignments)
4882.859 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRTNDGKITRPPGVK.....QKGRGSRKSPSPQPKKNV 1391

Scoring table:

BLOSUM62

Xgapop 10.0 , Ygapext 0.5

Xgapop 10.0 , Ygapext 0.5

Delop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n_model -DEV=x1h

-O=/gen2.1/USPTO_spool/US09512581/runat_23092003_163546_5464/app.query.fasta_1.1543

-DB=N_Geneseq_19Jun03 -OPMT=fastac -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0

-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum2 -TRANS=human40.cdi

-LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09512581.ecgn.1.1.0.0runat.23092003.163546.5464 -NCPU=6 -ICPU=3

-NO.MAP -LARGESOURCY -NEG.SCORES=0 -NAIT -DSPBLOCK=100 -LONGLOG

-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N_Geneseq_19Jun03:*

2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7193	100.0	5271	21 AAA28051	Human androgen shu
2	7193	100.0	5355	21 AAA28052	Human androgen shu
3	7155	99.5	7473	23 ABV22430	Human prostate exp
4	7149	99.4	7473	23 ABV28244	Human prostate exp
5	7148	99.4	7473	23 ABV25469	Human prostate exp
6	4760.5	66.2	6744	25 ABX14056	CDNA encoding huma
7	3047	42.4	5469	24 ABO99271	Human coding seque
8	2932.5	40.8	3957	21 AAA47423	Sequence encoding
9	1986.5	27.6	3920	23 ABL22103	Drosophila melanog
10	1729.5	24.0	6990	23 ABL22104	Drosophila melanog
11	1409.5	19.6	2496	24 ABK35358	Human cDNA encodin
12	967	13.4	1104	24 ABS56684	Euchromosome fragm
13	912.5	12.7	7943	22 AAK85948	Human immune/haema
14	864	12.0	714	22 AAS44727	Human full-length
15	762.5	10.6	1217	21 AAF18338	Lung cancer associ
16	748.5	10.4	2295	22 AAS02396	Human secreted pro
17	706.5	9.8	2212	22 AAH17132	Human cDNA sequenc
18	698	9.7	439	20 AAH87629	EST clone DY17. H
19	645	9.0	772	22 AAH03894	Human gene express
20	609	8.5	738	20 AA215259	Human cDNA clone (
21	591	8.2	461	24 ABL89669	Human polynucleoti
22	513	7.1	351	21 AA242861	Human 5' EST Isola
23	501	7.0	295	21 AAC02817	Human secreted pro
24	375.5	5.2	1489	22 AAH18066	Human cDNA sequenc
25	370	5.1	785	22 AAH07766	Human cDNA clone (
26	357	5.0	227	24 ABO94421	Tumour suppression
27	349	4.9	530	21 AA280598	Human colon cancer
28	276	3.8	433	22 AAS44899	Human cortig polyn
29	248.5	3.5	9361	23 ABV25498	Human prostate exp
30	235	3.3	20448	23 ABL19989	Human prostate exp
31	232	3.2	4493	25 ACA03930	Drosophila melanog
32	232	3.2	8527	22 ABL60039	CDNA downregulated
33	231.5	3.2	9274	24 ABL13224	Human polynucleoti
34	231	3.2	8503	24 ABL18253	DNA encoding novel
35	229	3.2	12777	23 ABL28641	Human polynucleoti
36	229	3.2	7992	24 ABL32895	Drosophila melanog
37	229	3.2	8083	23 AAS70692	DNA encoding human
38	225	3.2	7034	22 AAH19496	Human coding novel
39	225	3.1	5334	19 AAX14598	Murine coding sequ
40	223	3.1	557	24 ABL35631	H. pylori GHP0 175
41	222.5	3.1	11116	22 AAI59270	Human eosinophil-m
42	220.5	3.1	7713	22 AAK51958	Human polynucleoti
43	216.5	3.0	4534	23 ABV22145	Human polynucleoti
44	216.5	3.0	4534	23 ABV27985	Human prostate exp
45	216.5	3.0	19303	23 ABL22098	Human prostate exp

ALIGNMENTS

RESULT 1

ID AAA28051 standard; cDNA: 5271 BP.

AC AAA28051;

DT 01-DEC-2000 (first entry)

DE Human androgen shutoff gene 3 (AS3) cDNA sequence.

KW Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;

KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;

KW diagnosis; treatment; cytostatic; human; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

FT	exon	1..46
FT		/*tag= a
FT	exon	/number= 1
FT		47..173
FT		/*tag= b
FT	CDS	/number= 2
FT		66..4241
FT		/*tag= c
FT		/product= "AS3"
FT		/note= "nitrogen shutoff gene 3 protein, the CDS is
FT	exon	specifically claimed as SEQ ID #3"
FT		174..377
FT		/*tag= d
FT	exon	/number= 3
FT		378..464
FT		/*tag= e
FT	exon	/number= 4
FT		465..562
FT		/*tag= f
FT	exon	/number= 5
FT		563..689
FT		/*tag= e
FT	exon	/number= 6
FT		690..854
FT		/*tag= f
FT	exon	/number= 7
FT		855..911
FT		/*tag= g
FT	exon	/number= 8
FT		912..1027
FT		/*tag= h
FT	exon	/number= 9
FT		1028..1122
FT		/*tag= i
FT	exon	/number= 10
FT		1123..1268
FT		/*tag= j
FT	exon	/number= 11
FT		1269..1420
FT		/*tag= k
FT	exon	/number= 12
FT		1421..1534
FT		/*tag= l
FT	exon	/number= 13
FT		1535..1616
FT		/*tag= m
FT	exon	/number= 14
FT		1617..1665
FT		/*tag= n
FT	exon	/number= 15
FT		1666..1805
FT		/*tag= o
FT	exon	/number= 16
FT		1806..1921
FT		/*tag= p
FT	exon	/number= 17
FT		1922..2027
FT		/*tag= q
FT	exon	/number= 18
FT		2028..2188
FT		/*tag= r
FT	exon	/number= 19
FT		2189..2312
FT		/*tag= s
FT	exon	/number= 20
FT		2313..2471
FT		/*tag= t
FT	exon	/number= 21
FT		2472..2540
FT		/*tag= u
FT	exon	/number= 22
FT		2541..2677
FT		/*tag= v

CC	FT	/number= 23
CC	FT	2678..2801
CC	FT	/*tag= w
CC	FT	/number= 24
CC	FT	2802..3006
CC	FT	/*tag= x
CC	FT	/number= 25
CC	FT	3007..3121
CC	FT	/*tag= y
CC	FT	/number= 26
CC	FT	3122..3254
CC	FT	/*tag= z
CC	FT	/number= 27
CC	FT	3255..3374
CC	FT	/*tag= aa
CC	FT	/number= 28
CC	FT	3375..3437
CC	FT	/*tag= ab
CC	FT	/number= 29
CC	FT	3438..3583
CC	FT	/*tag= ac
CC	FT	/number= 30
CC	FT	3584..3689
CC	FT	/*tag= ad
CC	FT	/number= 31
CC	FT	3690..4129
CC	FT	/*tag= ae
CC	FT	/number= 32
CC	FT	4130..4354
CC	FT	/*tag= af
CC	FT	/number= 33
CC	FT	4320..4324
CC	FT	/*tag= ag
CC	FT	/note= "Destabilizing signal"
CC	FT	4355..5253
CC	FT	/*tag= ah
CC	FT	/number= 34
CC	FT	5228..5233
CC	FT	/*tag= ai
CC	FT	5249..5253
CC	FT	/*tag= aj
CC	FT	/note= "Poly adenylation cleavage site"
CC	XX	
CC	PX	WO2000050454-A1.
CC	PD	31-AUG-2000.
CC	XX	
CC	PF	24-FEB-2000; 2000OWO-US04732.
CC	XX	
CC	PR	24-FEB-1999; 99US-0121461.
CC	PA	(TUFT) TUFTS COLLEGE.
CC	XX	
CC	PI	Soto AM, Sonnenschein C, Geck P, Szelei J;
CC	DR	WPI: 2000-565451/52.
CC	XX	P-PSDB: AAY94702.
CC	PT	New human androgen-induced tumor suppressor cDNA sequence termed
CC	PT	'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
CC	PT	diagnosis and treatment of prostate cancer -
CC	XX	
CC	PS	Claim 1; Fig 1; 152pp: English.
CC	XX	
CC	CC	This invention relates to a human androgen-induced tumour suppressor cDNA
CC	CC	sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
CC	CC	on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
CC	CC	proliferation and use as a marker for the efficient diagnosis and
CC	CC	treatment of prostate cancer. The invention includes AS3 cDNA and protein
CC	CC	sequences, a vector comprising the cDNA sequence, a host cell transfected
CC	CC	with the expression vector, and a method for producing an AS3 polypeptide
CC	CC	comprising culturing the transfected cells. AS3 has cytostatic activity,
CC	CC	and acts to suppress cell proliferation. The AS3 gene is useful as a

CC marker for the efficient diagnosis and treatment of prostate cancer. The
CC AS3 nucleic acid molecule can be used as a source of antisense agents for
CC sequence specific modulation of gene expression. The AS3 protein may be
CC used in the treatment of disorders caused by aberrant modification or
CC mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
CC or aberrant post-translational modification of the AS3 protein. This
CC sequence represents the human AS3 cDNA sequence.

XX Sequence 5271 BP: 1782 A; 944 C; 1066 G; 1479 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	5271
Score:	7193.00	Matches:	1391
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-512-581b-2 (1-1391) x AAA28051 (1-5271)

```
QY 1 MetaIaHISerLysThrArGThrAsnAspGluLysIleThr-TyrProGluValLys 20
DB 66 ATGGCTCATCTCAAAACATGACCAATGATGGAAAAATTACATTCGGCTGGGTCAG 125
QY 21 GluIleSerAspLysIleSerLysGluMetValArgArgLeuLysMetValLys 40
DB 126 GAAATATCAGATAAATATCTAAGAGGAGATGGTACGATTAAGATGGTTGGAA 185
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
DB 186 ACTTTATGATGATGACGACGACGACTCTGMAAGAAAGACCTTATTAACCTAGCT 245
QY 61 LeuHISleuAlaSerAspPhePheLeuLysHISProGluLysAspValArgLeuVal 80
DB 246 TTACATCTTCTTCAGATTTTCTCTCAAGCATCTCGTAAAGATGTTGCTTACTGTA 305
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
DB 306 GCCTCGCTGCTTGCATATTTTTCAGGATTTATGCTCCGAAAGCTCCTTACACATCCCT 365
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuAspThr 120
DB 366 GATAAACAAGAGATATTTATGTTTATMACAAGACAGTGAAGGGCTTACAGGATACA 425
QY 121 LysSerProGlnPheAsnArgTyrPheTyrIleLeuGluAsnIleAlaIleTyrAllysSer 140
DB 426 AAGACCCACATTAATAGGATTTTATTTTACTTGAACATCTGGCTGCAAGTCA 485
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
DB 486 TATAACATATGCTTGGAGTTAGAACATAGCATGAATTTTCAACCAGCTATACAGAAC 545
QY 161 LeuPheSerValIleAsnAsnGlnIyHisAsnGlnLysValHisMetHisMetValAspLeu 180
DB 546 TTATTATTACGTTTAAACAATGGCCACATCAAGAAAGTCCATATGCATGCTAGACCTT 605
QY 181 MetSerSerIleIleCysGluGluLysPheThrValSerGlnGluLeuLeuAspThrValLeu 200
DB 606 ATGACCTCTATTATTGTTGAAAGGTATACAGTGTCTCAGAGACCTTTGGATACGGTTTAA 665
QY 201 ValAsnLeuValProAlaHISLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
DB 666 GTAATTCGTGTCACGTCATAGAATTTAAACAAGCAAGCATATGATTTGGCAAAAGCT 725
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
DB 726 TTACGAAAGAGCAGCAGCTCAACGCTATGAGCCATATATATACCATTTTATATCAGCTT 785
QY 241 LeuMetIleGluLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
DB 786 CTGATGCTTGGAAAAACATCTATCAGCCATTTGTCTCAGACATGCTCTTATCTTAATTTTG 845
QY 261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
```

```
DB 846 GAGCTCTACAAATATGATGATGATTCATTTGCTGCTCTGTTTACCCAGCTGAATTA 905
QY 281 LeuLysSerAsnAspAsnGluLysArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
DB 906 TTAAAGAGCAAGATAATAGAGAGCCCTACAGATGTTAACTACTGGCAAAAATGTTT 965
QY 301 GluAlaLysAspSerGluLeuValAserGlnAsnLysProLeuTProGlnCysTyrLeuGly 320
DB 966 GGGGCAAGAGATTTCAGAAATTTGGCTTCTCAAAACACCCCTTTGGCAGTGCTACTGGGC 1025
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
DB 1026 AGCTTTAAAGATATCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1085
QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
DB 1086 CTCATGAACCATTCCTGATTTTACCAAAAGACTTACAGAGATATCTTAAAGTGAAGTCACAT 1145
QY 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
DB 1146 GACCTCGAGGAAGCTATTTGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1205
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
DB 1206 GATATTCCTGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1265
QY 401 ArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysTyrAla 420
DB 1266 CGATGAGAGTACGCAAAAGAACCATGAGAGGACCTGCCAAATTTATTAAGAAATAGCT 1325
QY 421 LeuGlnSerAlaIleGlyLysAspAlaAlaLysGlnIleAlaTProIleLysAspLysLeu 440
DB 1326 TTACAGTCAGCAGCTGGAAAGATGCTGCAAAACAGATGCAATGCAATCAAAACAAATGG 1385
QY 441 LeuHISleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
DB 1386 CTACATATATATATCAAAATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1445
QY 461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
DB 1446 CATATCATGTTCTCCACAAATTTAGAAATACAGAAACGATGAAATGCTTATATTAATCTTG 1505
QY 481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTyrLysCysGlnAsn 500
DB 1506 TATGCCACACTGATTTAAATGCTGTGAAGACATTTGAATGAATGGGAAATGTCAAAT 1565
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
DB 1566 CTGCTCCGACATCAAGTAAAGATTTGCTTGACTTGATTAACCAACCCAAACAGATGCC 1625
QY 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
DB 1626 AGTGTCAAGGCATATTTTCAAAAGGATTAATTAACAAGAAATTTACCTATCTCTGCT 1685
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg 560
DB 1686 AAGGCTCAGAGATTTCAATGAGAAATTCACACAGGTTTGAAGATGATATAGAAATTAAGA 1745
QY 561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluLysCysVal 580
DB 1746 AAGCAGTTAGAGTACTTGTAGTCCAAACATGCTCTGCAACAGAGCTGAAGTGTGTG 1805
QY 581 ArgGluIleThrLysLysLeuGluLysAspProLysGlnProThrAsnProPheLeuGluMet 600
DB 1806 CGTGAATATACCTAAGAGTTGGCAACCCCAACACCTTACAAATCTCTTCTGGAAGG 1865
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
DB 1866 ATCAAGTTTCTTGGAGAGGATAGCACTGTGACATATGATACCAATCTATCACTGCT 1925
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluVal 640
```

Dh 1926 CTATTAAACAGTGAACAAATCAATAGATGGAACACAGATGATGAAGATGAGGTGTT 1985
Qy 641 ProthraspGlnAlaIleArgAlaGlyLeuGlnLeuLeuLeuValLeuSerPheThrHis 660
Dh 1986 CCAACTATATACAGCATCAGACGAGGTCTTGAACTGCTTAAGTACTCTCTATTACACAT 2045
Qy 661 ProIleSerPheHisSerAlaGluThrPheGlnSerLeuLeuAlaCysLeuLysMetAsp 680
Dh 2046 CCATCTCATCTTCTCTGCGAAGACATTGAAATCATTACGCTGCTGCTGAATAATGAT 2105
Qy 681 AspGlnLysValAlaGlnAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Dh 2106 GATGAAAAAGTAGAGAGAGCTGCACTCAAAATTTTCAAAAAACAGAGCAAAATTTGAA 2165
Qy 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLysLys 720
Dh 2166 GAGATTTTCCACATCAGATCAGCTTGCTTCTGCTTTTACATCCAAATCTAAAAA 2225
Qy 721 GlyProAlaArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
Dh 2226 GGAACCCCCCGTCAGCCAAATATGCCATTCATGTATCCATGCGATATTTTCTAATAA 2285
Qy 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisIleLysSerLysAspProSerAsnLeu 760
Dh 2286 GAGACCCAGTTTGACAGATATTTGAGCTCTGCAATAGAGCCCTAGATCCAAAGAACCTG 2345
Qy 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuAlaProAspGln 780
Dh 2346 GAACATCTCATTAACACCATTTGGTACTATTGGTCATATTGCTCTCTGCGACGATCA 2405
Qy 781 PheAlaAlaProThrLysSerThrProValAlaThrPheIleValLysAspLeuLeuMetAsn 800
Dh 2406 TTTCGCTCTCTTGGAAATCTTGGTAGCTAATTTCATTTGGAAGAATCTTCTCATGAT 2465
Qy 801 AspArgLeuProGlyLysLysThrThrLysLeuThrProValProAspGlnGluValSerPro 820
Dh 2466 GATCGGCTTCAGGAGAAAAAGACACAACTTTGGGCTCCAGATGAAAGATGATCTCCT 2525
Qy 821 GluThrMetValLysIleGlnAlaIleLysMetLeuValArgThrPleuLeuGlyMetLys 840
Dh 2526 GAGCAATATGTCAAAAATTCAGGCTATTAATAATGATGGTTCGATGGCTACTTGGATGAA 2585
Qy 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
Dh 2586 AATATATACAGTAAATACAGAGAACTTCTAACCTTAAGATTGCTAACACAAATTTCCATAGT 2645
Qy 861 AspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Dh 2646 GATGAGACTTGACAGAACAGGGAATATTAGTAAACCAGATATGTCACTGTCAGACTT 2705
Qy 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnIleProCysTyrHisGluIleThr 900
Dh 2706 GCTCTCTGGAGTCTATTGTGTAAGCTGGACAAAGACCTGTACCATGAAATCATCA 2765
Qy 901 LeuGlnIleThrGlnLeuLysValAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Dh 2766 TTGAACAATAATACGCTATGTGCAATAGCTATACAGATGATCTATCAAGTAAAGAA 2825
Qy 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
Dh 2826 GTGTTTGCCCAAGAACTTCACAAAGGCTTCCCGCTTTACGCGCTCCACTTGAGTATATG 2885
Qy 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Dh 2886 GCAATCTGTGCGCTTTGCGAAAAAGATCCTGTAAAGAGAGAGAGCGTCATGAGGAA 2945
Qy 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
Dh 2946 TGTTTGGTGAATAATATATATAGTAAAGCGGAGATATCTGAAGACGATGCGAGCTTTAGT 3005
Qy 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Dh 3006 GAAAAAATTAATTTCTTTCTAACAGAGTATGTGTTCATTAATAATTCACCTTTTGGCA 3065

Qy 1001 HisAspProAspTyrValLysValGlnAspIleGlnIleLeuLysAspValLysGluCys 1020
Dh 3066 CATACCCACATTTATGTCMAAGTACAGATATTTCACAACTTAAGATGTTAAAGATGT 3125
Qy 1021 LeuThrPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Dh 3126 CTTTGGTTTGTTCGAAATATTAATGCTTAATAATGATAATACATGCAAGCTTTTATC 3185
Qy 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Dh 3186 AGAAGATGTGAGAAATATTAAACAAACAAAGATGCCCAAGAGCAGATGATGCAAA 3245
Qy 1061 MetAsnGluLysLeuTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSer 1080
Dh 3246 ATGAATGAAAAACGTACAGTGTGTGATGTTCCATGAATATCATCATGTCAAAGAT 3305
Qy 1081 ThrThrTyrSerLeuGlnSerProLysAspProValLeuProAlaArgPheThrGln 1100
Dh 3306 ACTACATACAGTTTGGAAATCTCTTAAGACCCGGTACTACAGCTCTTCTTCACCTCAA 3365
Qy 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
Dh 3366 CCTGACAAAGATTTCACTAACACCAAAATATATCTGCTCTGAATGAATCATTTTTC 3425
Qy 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Dh 3426 ACTCCTGAAAAACCTAAACAAACCAATGTTCTAGACCTGTAAACAGCCACTTTCATCA 3485
Qy 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Dh 3486 GCAGGCAAGCATATCTGACCAAAATCAATCAAGAAATGGAACCTGTAACCAATGCAAGAGC 3545
Qy 1161 SerSerAsnProSerSerProGlyArgIleLysGlyLysArgLeuAspSerSerGluMetAsp 1180
Dh 3546 AGCTCAATCCAACTCTCTCGAAGAATAAAGGGAGCGTGTATTTCTGAAATGAT 3605
Qy 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Dh 3606 CACAGTGAATAATGAAATGAAATTCACAAATCTTCACTTGGCGGGAATAAAAGTCAAG 3665
Qy 1201 ArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArgGlyArgLysLysThr 1220
Dh 3666 AGACACACTGTGATTTCTGTAGGCTCTGAATTTGAGAGACCTACAGGCAGAAAAAACG 3725
Qy 1221 ProValThrGlnGlnGlnLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
Dh 3726 CCCGTCAAGAAAGAGAGGAAATTAAGTATGATGACTGACTGAATGTTGCTACAGAA 3785
Qy 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
Dh 3786 CAGAAACCTAAAGCAGTACAGCAAGTCCGAAAAAGAGCCCTACAGGCTTCAGAAATCGAT 3845
Qy 1261 GlnGlnIleThrProGlnGlnLysArgLeuLysGlnAspIleLeuGlnIleGlnLysGln 1280
Dh 3846 GAACAGCAGTGGCTTGAGGAAAGAGGCTCAAAAGATATATTAGAAATGAAGAGTGA 3905
Qy 1281 GlnAsnSerProProLysLysGlyLysArgGlyLysArgProProLysProLeuGlyGly 1300
Dh 3906 CAGATATGTCGCCCAAAAAAGGTAAAGAGCGGACACCAAAACCTCTTGCGAGAGT 3965
Qy 1301 ThrProLysGlnGluProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
Dh 3966 ACACCAAAAGAGAGCCACAAATGAATACTCTAAAAAGAGAACAAAAAATAATCTGGA 4025
Qy 1321 ProProAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1340
Dh 4026 CCTCCACACCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4085
Qy 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGlu 1360
Dh 4086 TCCAAAAGCCAAACAGCACCGAGTGTCAAGGAGACACGACGAGAGAGAGATTTCTCTGAA 4145

QY	1361	SeqSerAaAllegluserThnGlnSerThrProGlnLysGlyArgGlyArgProSerLys	1360
QY	1361	SeqSerAaAllegluserThnGlnSerThrProGlnLysGlyArgGlyArgProSerLys	1360
Db	4146	TCCTAGTGGAAATGTGAATCCACACAGTCCACACACAGAAAGGACGAGGAACCATCAAAA	4205
OY	1381	ThrProSerProSerGlnProLysLysAsnVal	1391
Db	4206	ACGCCATTCACCATCACAACCAAAAAAATATGTG	4238
RESULT 2			
AC	AAA28052	standard; cDNA; 5355 BP.	
XX	AAA28052;		
XX	01-DEC-2000	(first entry)	
DE	Human androgen shutoff gene 3 (AS3)	CDNA sequence SEQ ID #4.	
XX			
KW	Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;		
KW	Chromosome 13q12-13q: cell proliferation inhibitor; prostate cancer;		
KX	diagnosis; treatment; cytostatic; human; ss.		
XX			
OS	Homo sapIens.		
Key	Location/Qualifiers		
FT	85..130		
FT	/*tag= a		
FT	/number=1		
FT	131..257		
FT	/*tag= b		
FT	/number=2		
FT	150..4325		
FT	CDS		
FT	/*tag= c		
FT	/product= "AS3"		
FT	/note= "Androgen shutoff gene 3 protein"		
FT	238..461		
FT	/*tag= d		
FT	/number= 3		
FT	462..548		
FT	/*tag= e		
FT	/number=4		
FT	549..646		
FT	/*tag= f		
FT	/number= 5		
FT	647..773		
FT	/*tag= e		
FT	/number= 6		
FT	774..938		
FT	/*tag= f		
FT	/number= 7		
FT	939..995		
FT	/*tag= g		
FT	/number= 8		
FT	996..1111		
FT	/*tag= h		
FT	/number= 9		
FT	1112..1206		
FT	/*tag= i		
FT	/number= 10		
FT	1207..1352		
FT	/*tag= j		
FT	/number= 11		
FT	1353..1504		
FT	/*tag= k		
FT	/number= 12		
FT	1505..1618		
FT	/*tag= l		
FT	/number= 13		
FT	1619..1700		
FT	/*tag= m		
FT	/number= 14		
FT	1701..1749		
FT	/*tag= n		
FT	exon		

FT		/number= 15
FT	exon	1750..1889
FT		/*tag= o
FT		/number= 16
FT	exon	1890..2005
FT		/*tag= p
FT	exon	/number= 17
FT		2006..2111
FT		/*tag= q
FT	exon	/number= 18
FT		2112..2272
FT		/*tag= r
FT	exon	/number= 19
FT		2273..2396
FT		/*tag= s
FT	exon	/number= 20
FT		2397..2555
FT		/*tag= t
FT	exon	/number= 21
FT		2556..2624
FT		/*tag= u
FT	exon	/number= 22
FT		2625..2761
FT		/*tag= v
FT	exon	/number= 23
FT		2672..2885
FT		/*tag= w
FT	exon	/number= 24
FT		2886..3090
FT		/*tag= x
FT	exon	/number= 25
FT		3091..3205
FT		/*tag= y
FT	exon	/number= 26
FT		3206..3338
FT		/*tag= z
FT	exon	/number= 27
FT		3339..3458
FT		/*tag= aa
FT	exon	/number= 28
FT		3459..3521
FT		/*tag= ab
FT	exon	/number= 29
FT		3522..3667
FT		/*tag= ac
FT	exon	/number= 30
FT		3668..3773
FT		/*tag= ad
FT	exon	/number= 31
FT		3774..4213
FT		/*tag= ae
FT	exon	/number= 32
FT		4214..4438
FT		/*tag= af
FT	exon	/number= 33
FT		4314..4408
FT		/*tag= ag
FT	misc_signal	/note= "Destabilizing signal"
FT		4439..5337
FT	exon	/*tag= ah
FT		/number= 34
FT	polyA_signal	5312..5317
FT		/*tag= ai
FT	misc_feature	5333..5337
FT		/*tag= aj
FT		/note= "Poly adenylation cleavage site"
PN	WO20050454-A1.	
XX		
PD	31-AUG-2000.	
XX		
PF	24-FEB-2000; 2000WO-US04732.	
XX		

PR 24-FEB-1999; 9905-0121461.
 XX (TUFT) TUFTS COLLEGE.
 PA Soto AM, Sonnenschein C, Geck P, Szelei J;
 PI WPI; 2000-565451/52.
 XX DR P-PSDB; AAY94702.
 XX
 PT New human androgen-induced tumor suppressor cDNA sequence termed
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
 PT diagnosis and treatment of prostate cancer -
 XX
 XX Example 4; Fig 6; 152pp; English.
 PS
 XX This invention relates to a human androgen-induced tumour suppressor
 CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
 CC located on chromosome 13 at position 13q12-13q. AS3 has a role in
 CC inhibiting cell proliferation and use as a marker for the efficient
 CC diagnosis and treatment of prostate cancer. The invention includes AS3
 CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host
 CC cell transfected with the expression vector, and a method for producing
 CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
 CC cytosolic activity, and acts to suppress cell proliferation. The AS3
 CC gene is useful as a marker for the efficient diagnosis and treatment of
 CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
 CC antisense agents for sequence specific modulation of gene expression. The
 CC AS3 protein may be used in the treatment of disorders caused by aberrant
 CC modification or mutation of a gene encoding an AS3 protein, misregulation
 CC of the AS3 gene or aberrant post-translational modification of the AS3
 CC protein. This sequence represents the human AS3 cDNA sequence with an
 CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
 CC compared with the claimed AS3 cDNA sequence AAA28051.
 CC
 XX
 SQ Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 5355
 Score: 7193.00 Matches: 1391
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-512-581b-2 (1-1391) x AAA28052 (1-5355)
 QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
 DB 150 ATGGCTCATTCAAAGACTAGGACCAATGATGGAATAATTCATATCCGCGCTGGGGTCAAG 209
 QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
 DB 210 GAATATCATGATAAATATCTAAAGAGAGATGCTGAGCATTAAGATGCTGTGTA 269
 QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
 DB 270 ACTTTTGTGATATGACACGAGACTCTGAAGAAAGAAAGAGCTTTTAAACCTAGCT 329
 QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
 DB 330 TTACATCTTGCTTCAGATTTTCTTCACAGCATCTGTTAAAGATGCTTACTGTA 389
 QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
 DB 390 GCGTGGCTGGCTTGATATTTTCAGATTTATGCTCTGAGCTCTTACACATCCCT 449
 QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
 DB 450 GATTAACCTAAAGCTATATTTATCTTATTAACAAGACAGTTGAAGGGGCTAGAGATACA 509
 QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTTPValLysSer 140
 DB 510 AAGAGCCCAATTCATAGGATTTTATTTACTTGAAGACATTGCTGGGTCAAGTCA 569

QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
 DB 570 TATACATATGCTTGTTGATAGAGATAGCAATGAATTTTCCACCACGATACAGAAC 629
 QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
 DB 630 TTATTTTCACTTAATAAACAATGGCCCAACATCAAGAAAGTCCATATGACATGATGACCTT 689
 QY 181 MetSerSerIleIleCysGluGlyAspThrValSerGlnLeuLeuAspThrValLeu 200
 DB 690 ATGAGCTCTATTTATTTGTGAAGGATGATACAGTGTCTCAGGAGCTTTTGGATAGGTTTA 749
 QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
 DB 750 GTAATGTGGTACTCGCTCATTAAGAAATTTAAACAGACACATATGATTTGGCAAGGCT 809
 QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
 DB 810 TTACTGAAGAGGACAGCTCAAGCTATGAGCCATATATTCACACTTTTAAATCAGGTT 869
 QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
 DB 870 CTGATGCTTGGAAGAACATCTATCAGCGATTTGTACAGCAGATGCTTTGACTTAATTTTG 929
 QY 261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
 DB 930 GAGCTTACATATTTGATGATGATCATTTCTCTCTCTCTGTTTACCCAGCTGAATTTAA 989
 QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
 DB 990 TTAAGAGCAATGATTAATGAGAGCGCTTCAAGTTGTTAACTACTGCGCAAAAAGTTT 1049
 QY 301 GlnValAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGly 320
 DB 1050 GGGCAAGAGATTCAGAAATTTGCTTCTCAAAACAGCACCTTTGGCAGCTCTACTTGCC 1109
 QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
 DB 1110 AGTTTAAATGATATTCATATTCATACCAATCCGCTGGAATGTGTGAATTTGCTACCATTTGT 1169
 QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
 DB 1170 CTCATGACCATCTCGATTTAGCAAAAGACTTAAACAGATATCTTAAAGGAGGCTCAT 1229
 QY 361 AspProGluGlnAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
 DB 1230 GACCTTAGAGAGCTATTAACATGATGATTTTGTGCAATAGTTACAGCTGCTAAAG 1289
 QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
 DB 1290 GATATCTCTGGTCATGATGATCACTTAAATTTTGTGAGAGAGAACATTAAGACAA 1349
 QY 401 ArgThrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysTyrAla 420
 DB 1350 CGATGAGAGCTACCCAAAGAACCATGATGGAGCTTCCCAAAATTTTAAAGAAATATGCT 1409
 QY 421 LeuGlnSerAlaIleGlyLysAspAlaAlaLysGlnIleAlaTyrIleLysAspLysLeu 440
 DB 1410 TTACAGTCACAGCTGGAAGAAAGATGCTGCCAAACAGATACATGATGATCAAGCAATTTG 1469
 QY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
 DB 1470 CTACATATATATATCAAAATAGATATGATGATGATGATGATGATGATGATGATGATGAT 1529
 QY 461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
 DB 1530 CAATACATGTGCTCCCAATTTTGAAGAACACAGAAAGGATGAAGTCTTATATTTACTTGG 1589
 QY 481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTyrPlyLysCysGlnAsn 500
 DB 1590 TATGCCACTGATTTTAAATGCTGTGTGAAGCATTTGAATGTAATGTGGAATGTCAAAAT 1649

QY	501	LeuLeuAaYghISgInValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla	520
Db	1650	CGGCGCCGCAATCAAGTAAGATTTGGCTTGACTTGATTAAGCAACCCAAAACAGATGCC	1709
QY	521	SetValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly	540
Db	1710	ACTGTCAGAGCCCATTTTTCAAAAGAGTGGTTATTACAGAAATTTTACCTGATCTCGGT	1769
QY	541	LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg	560
Db	1770	AAGCGTCAGAGATTTCATGAGAATTCACACAGGGTTTGAAGAATGATGAGAAATTAACA	1829
QY	561	LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluLysCysVal	580
Db	1830	AAGCGATTAGAAGTACTTGTTAGTCCAACATGCTCTCGCAACGAGCTGAAGGTGTGGTG	1889
QY	581	ArgGluIleThrLysLeuLeuGlnLysAsnProLysGlnProThrAsnProPheLeuGluMet	600
Db	1890	CGTGAATTAACCTTAAGAAAGTTGGGCAACCCCAACAGCCTTACAAATCTTTTCCGTGAAATG	1949
QY	601	IleLysPheLeuLeuGluIleArgIleLeuAlaProValHisIleAspThrGluSerIleSerAla	620
Db	1950	ATCAAGTTTCTCTTGGAGAGATGACCTGACCTGGCACAATGATGACCGAATCTTATCGAGTCT	2009
QY	621	LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluLysVal	640
Db	2010	CTTATTATAACAAGTATACAAATTCATATGATGAGACAGACAGATGATGAAGATGAGGGTCTT	2069
QY	641	ProThrArgAsnAlaIleArgAlaGlyLeuGluIleuLeuLysValLeuLeuSerPheThrHis	660
Db	2070	CCAACTGATCAAGCCATACAGACAGAGCTCTTGACATGGCTTAAAGTACTCTCATTTACACAT	2129
QY	661	ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp	680
Db	2130	CCCACTCATTTTCATTTCTGCTGAACATTGTAATCATTAATCTGCTGCTGCTGAATATGAT	2189
QY	681	AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu	700
Db	2190	GATGAAAAAGAGCAAGAACTGCATCAACAAATTTTCAAAACACAGGAAGCAAAATTTGAA	2249
QY	701	GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys	720
Db	2250	GAGGATTTTCCACACATACAGATCAAGCTTGGCTCTGCTGTTTTCATCAACAAATCTAAAAA	2309
QY	721	GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys	740
Db	2310	GGACCCCCCGCAAGCAAAATATGCCATTCATTGTATCCATGCCATATTTTCTAGTAAA	2369
QY	741	GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu	760
Db	2370	GAGACCCCGTTGGACAGATATTTTGGAGCCCTGCTGCATAGAGCCTTGATCCAAAGCAACCTG	2429
QY	761	GluHisIleuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln	780
Db	2430	GAACTCTCATTAACCCATTTGGTATCAATTTGGTCAATTTGGCTCTTGGCACCTGATCAA	2489
QY	781	PheAlaAlaProTyrLysSerTyrValAlaThrPheIleValLysAspLeuMetAsn	800
Db	2490	TTTGCTGCTCTTGGAAATCTTTGGGTACTTCTTCATTTGTAAGAAAGATCTTCTCATGAT	2549
QY	801	AspArgLeuProGlyLysLysThrThrLysLeuTyrValProAspGluGluValSerPro	820
Db	2550	GATCGGCTCCAGAGGAAAAACAACATTAACCTTGGGTCCAGATGAAGAATATCTCTCT	2609
QY	821	GluThrMetValLysIleGlnAlaIleLysMetMetValArgTyrPheLeuGlyMetLys	840
Db	2610	GAGACAAATGGTCAAAATTCAGGCTATTAATGATGTTGCATGGCTACCTTGGATGAAA	2669
QY	841	AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer	860
Db	2670	AATAATACATTAATACAGAACTTCTACCTTAAAGATTGCTTAACACACATATTTGCTATGCT	2729
QY	861	AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu	880
Db	2730	GATGGAGACTTGACAGAAACAGGGAAAAATTAGTAAACCAGATATGTCACGCTGAGACTT	2789
QY	881	AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr	900
Db	2790	CGCTGTGGAGTGGCTATTGTGTAAGCTGGCACAAAGAACCTGTGTACCAATGAATATATACA	2849
QY	901	LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln	920
Db	2850	TTAGACAAATATACAGTATGTGTCATTTAGCTATCAACGATGAAATGCTATCAAGTAAGCAA	2909
QY	921	ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet	940
Db	2910	GTTGTTGGCCCAAAACCTTCACAAAGGCCCTTTCCGTTTACGGCTTCCACCTGAGTAAATG	2969
QY	941	AlaIleCysAlaLeuCysAlaLysAspProValLysGluIleArgAlaHisAlaArgGln	960
Db	2970	GCAATCTGTGCCCTTTGTGCAAAAGATCTCTTAAGGAGAAAGACGCTCATGTAGGCCAA	3029
QY	961	CysLeuValLysAsnIleAsnValArgArgIleuTyrLeuLysGlnHisAlaIleValSer	980
Db	3030	TGTTTGGTGAATAATATTAATGTAAAGGGGAGATGTGAAAGCAGCATGCGAGCTTATG	3089
QY	981	GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla	1000
Db	3090	GAAAAATTATGTCTCTCTCAACAGATATGTTTCCATATACAAATTCACCTTTTGCA	3149
QY	1001	HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys	1020
Db	3150	CATGACCCAGATATATGTCAAAAGTACAGGATATTAACACACTTAAAGATGTTAAAGAAATGT	3209
QY	1021	LeuTyrPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle	1040
Db	3210	CTTTGGTTGTTCGAAATATTAATGAGCTTAATAATGAATAATACAGTACCGCTTTATTC	3269
QY	1041	ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys	1060
Db	3270	AGAAAGATGTGTAAGAAATATTAAACAAACAAACAAATCCCAAGGACCGATGATGACAAA	3329
QY	1061	MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer	1080
Db	3330	ATGATGAAAAACTGTACACTGCTGTGTGATGTTGCCATGAATATCATCATGTCAAAGACT	3389
QY	1081	ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln	1100
Db	3390	ACTGATATACATTTTGGAAATCTCCTTAAGAGCCGGTACTACAGCTGCTTCTTCACGCA	3449
QY	1101	ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe	1120
Db	3450	CCTGACAAAGATTTTCAGTAACACCAAAATATATTCCTGCTCGTGAAGAAATCATCTTTTC	3509
QY	1121	ThrProGlyLysProLysThrThrAsnValLeuGlnValAlaValAsnLysProLeuSerSer	1140

Db 3810 CCCGTCACAGACAGGAGGAAATTAGCATGATACCTTACTAAAGTTGGTACAGAA 3869
Qy 1241 GlnLysProLysGlySerGlnArgSerArgLysArgLysIleThrAlaSerGluSerAsp 1260
Db 3870 CAGAAACCTTAAGGCACTACGCAAGTCGGAAGAGCCATACGGCTTCAGATCTGAT 3929
Qy 1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
Db 3930 GAACAGCAGTGGCTGTGAGAAAGAGGCTCAAGAGATATATATGAAATGAAAGATGAA 3989
Qy 1281 GlnAsnSerProProLysGlyLysArgGlyArgProProLysProLeuGlyGly 1300
Db 3990 CAGATATGTCGCCCAAAAGAGGTAAAGAGCCGACCAACCACTCTTGGTGAGGT 4049
Qy 1301 ThrProLysGluGluProThrMetLysThrSerLysGlySerLysLysSerGly 1320
Db 4050 ACACCAAAAGAGAGCCACACATGAAACTCTTAAAAAGAACCAAAAAAATCTGGA 4109
Qy 1321 ProProLysProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
Db 4110 CCTCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4169
Qy 1341 SerLysSerLysGlnHisArgValSerArgArgArgGlnGlnGlnGlnGlnGlnGln 1360
Db 4170 TCCAAAAGCAACAGCACCGAGTGTCAAGAGACACAGACAGAGAGAGAGAGAGAGAG 4229
Qy 1361 SerSerAlaIleGlnSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db 4230 TCTAGTGCATTTGAAATCCACACAGTCCACACACAGAAAGAGACAGAGAGAGAGAG 4289
Qy 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4290 ACGCATCACCATCATCAACCAAAAAAATGTG 4322
RESULT 3
ABV22430
ID ABV22430 standard; cDNA: 7473 BP.
XX
XX ABV22430;
XX
XX 13-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 22421.
XX
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX
XX 16-MAR-2000; 2000US-189862P.
XX
XX 25-MAY-2000; 2000US-207454P.
XX
XX 09-JUN-2000; 2000US-211314P.
XX
XX 18-JUL-2000; 2000US-219007P.
XX
XX 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer -
PS Claim 1; Page 3901-3902; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV2213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 7473 BP; 2458 A; 1243 C; 1480 G; 2285 T; 7 other;

Alignment Scores:

Score: 0 Length: 7473
Pred. No.: 7155.00 Matches: 1387

Percent Similarity: 99.71% Conservative: 0

Best Local Similarity: 99.71% Mismatches: 4

Query Match: 99.47% Indels: 0

DB: 23 Gaps: 0

US-09-512-581B-2 (1-1391) x ABV22430 (1-7473)

Qy 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrYrProProGlyValLys 20
Db 158 ATGGCTCATTCAAAGACTAGACCAATGATGAGAAATTCACATTCGCGCTGGGCTCAAG 217
Qy 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
Db 218 GAAATATCAGATTAATAATATCTAAAGAGAGATGGTGGACGATTTAAAGATGGTTGAAA 277
Qy 41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuYrLeuAsnLeuAla 60
Db 278 ACTTTTATGATATGACACGAGACTCGAAGAAAGAGAGCTTTATTAACCTGAGCT 337
Qy 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
Db 338 TTACATCTTCCTTCAGATTTTCTTCACACATCCATTAAGAGTTCGCTTACCTGAGTA 397
Qy 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProYrThrSerPro 100
Db 398 GCGTGGCTGCTGCTGATATTTTCAGAGATTTATGCTGAGAGTCTTACACATCCCT 457
Qy 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
Db 458 GATTAACCTAAAGCATATATTTATTTATTAACAAGACAGTGAAGGGGCTAAGAGATACA 517
Qy 121 LysSerProGlnPheAsnArgYrPheYrLeuLeuGlnLysIleAlaIleTyrValLysSer 140
Db 518 AAGAGCCACACATTCATAGGTATTTTATTTACTTGAAGACATGCTTGGGCTCAAGTCA 577
Qy 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuYrArgThr 160
Db 578 TATTAACATATGCTTTGAGTTAAGAAGATAGCATTAATTTTCACCCGCTATACAGAAC 637
Qy 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
Db 638 TTATTTTCAGTTATTAACAATAGGCCACATACAGAAATCCATATGACATGATGAGACTT 697
Qy 181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
Db 698 ATGAGCTCTATTTATTTTGAAGGTGATACAGTGTCTGAGAGCTTTTGGATACGCTTTTA 757
Qy 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
Db 758 GTAAATCTGCTGCTCATAGCAATTTTAAACAGACAGACATATGATTTGGCAAGGCT 817

Qy 221 LeuLeuysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
|||||
Db 818 TTACTGAAGAGACAGCTCAAGCTATTGAGCCATATATATACCAATTTTTTTATATCAGGTT 877
Qy 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
|||||
Db 878 CTGATGCTGGGAAACATCTATACAGCATTTGTCAGAGCATGCTTTGACTTTAATTTTG 937
Qy 261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
|||||
Db 938 GAGCTTACCAATATGTGATGATCATTTGCTGCTCTCTGTTTACCCACCTTGAATTTTAA 997
Qy 281 LeuysSerAsnAspAsnGluGluArgLeuGlnValLysLeuLeuAlaLysMetPhe 300
|||||
Db 998 TTTAAGACCATGTAATGAGAGCGCTTACAACTGTTTAACTACTGCGCAAAATGTTT 1057
Qy 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGly 320
1058 GGGCCAAAGCATTCAGAAATTGCTTCCAAACAAAGCCACTTGGCAGTGGCTAGTGGGC 1117
Qy 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
|||||
Db 1118 AGGTTTATGATATCTCATGTAACCAATCCCTGGAAAGTGTGAAATTTGCTAGGCCATTTGT 1177
Qy 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
|||||
Db 1178 CTCATGAACCATCTCGATTTAGCAAAAGACTTAACAGAGTACTTTAAAGTAGGCAAT 1237
Qy 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
1238 GACCCCTAGGAAAGCTATTAGCATGATGTTATTGTCATATGATACAGCTGTAAAG 1297
Qy 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
1298 GATATTTCTTCTGCGCATGATGACTTACTTAATTTTGTGAGAGAGAGACATTTGACAAA 1357
Qy 401 ArgTrrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
1358 CGATGGAGAGTACGCAAAAGAACCATGATGGGACTGTGCCAAATTTATAAACAAATATGCT 1417
Qy 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrrPileLysAspLysLeu 440
1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATGATGATCAAAAGACAAATTG 1477
Qy 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
1478 CTACATATATATATATCAAAATAGATGATGATGATGATGATGATGATGATGATGATGCT 1537
Qy 461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
1538 CAATACATGCTTCTCTCAAAATTTAGAAACCTACAGACAGATGAAATGCTTATATTTACTTG 1597
Qy 481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrrPlyCysGlnAsn 500
1598 TATCCCACTGATGTTTAAATGCTGTCGAAACATTTGATGAAATGTCGAAATGTCAAAT 1657
Qy 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
1658 CTGCTCCGACATTCAGTAAGAAAGATTTGCTGATGATTAAGCAACCCAAACAGATGCC 1717
Qy 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
1718 AGTCTCAAGGCCATATTTTCAAAAGTATGTTATTTACAAAGAAATTTTACTGATCTTGCT 1777
Qy 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg 560
1778 AAGCTCAGATTTTCATGAAGAAATTCACACAGGCTGTAGAAGATGATGAGAAATAAGA 1837
Qy 561 LysLysLeuGluValLysLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
1838 AAGGAGTTAGAAAGTATGTTAGTTCACAAATGCTTCCAGACAGGCTGAAGGTTGCTTG 1897
Qy 581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGlnMet 600

Db 1898 CGTGAATATACTAGAAATTTGGCACCCCAACAGCTACAAATCTCTTCTCGAATG 1957
Qy 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
1958 ATCAAGTTTCTTGGAGAGGATAGCACTTGCACATAGATACCAATCTATCATAGTCT 2017
Qy 621 LeuIleLysGlnValAsnLysSerIleAspLysThrAlaAspAspGluAspGluGlyVal 640
2018 CTATTTAAACAAGTAACAAATCAATGATGAAACAGCAGATGATGAAGATGAGGCTGT 2077
Qy 641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
2078 CCAACTGATCAAGCATCAGACAGCTTCAACTGCTTAAGTACTCTCATTTTACACAT 2137
Qy 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
2138 CCCATCTCATTTCTATTTGCTGTAACAAATTTGAAATCATATAGCTTGTCTGAAATGAT 2197
Qy 681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerIleGlu 700
2198 GATGAAAAAGTAGAAGAAAGCTGCACATCAAAATTTTCAAAACACAGGAAGCAAAATGAA 2257
Qy 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
2258 GAGGATTTTCCACATCATGATCAGCTTGTCTCTCTGTTTACATCAACAAATCTTAAAAA 2317
Qy 721 GlyProProAlaGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
2318 GGACCCCCCTCACAAGCCAAATATGCCATTCATGATGATCAATGCGATTTTTCATGTA 2377
Qy 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspSerAsnLeu 760
2378 GAGACCCAGTTTGCACAGATATTTGAGCTTGCATAGAACCTTGAATCCAAAGCACTTG 2437
Qy 761 GluHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuLeuAlaProAspGln 780
2438 GACATCTCATTAACACCATGCTGTTACTATGTCATATGCTGCTGCTGCTGCTGCTGCTGCT 2497
Qy 781 PheAlaIleProTrrPlySerTrrPvalAlaThrPheIleValLysAspLeuMetAsn 800
2498 TTTGCTGCTCTTGAATCTTTTGGTAGCTCTTTCATATTTGAAAGATCTTCTCATGAT 2557
Qy 801 AspArgLeuProGluLysLysThrThrLysLeuTrrPvalProAspGluGluValSerPro 820
2558 GATGGCTTCCAGGGAAGAAACAACTAAACTTTGGTCCAGATGAAGAAATCTCTCT 2617
Qy 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrrPleuLeuGlyMetLys 840
2618 GAGACATGCTCAAAATTCAGGCTATTAAATGATGATGCTGATGCTGATGCTGATGATAA 2677
Qy 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
2678 AATATATCACAATAATATGAGAACTTCTTACCTTAAGATGTGAACAAATATTTCCATAGT 2737
Qy 861 AspGlyAspLeuThrGlnGluGlnLysIleSerLysProAspMetSerArgLeuArgLeu 880
2738 GATGAGACCTTGACAGAAACAGGGAAGAAATTAAGTAACACAGATATGCTACGCTGACGCT 2797
Qy 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnLysProCysTyrGlnHisGlnIleIleThr 900
2798 GCTGCTGAGACTGCTATTTGTAACTGTCGACACAAACACCTTGTACATGAATATGATACA 2857
Qy 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
2858 TTAGAACAATATCAGTATGATGCAATGATGATGATGATGATGATGATGATGATGATGAT 2917
Qy 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
2918 GTGTTTGCCGAAACCTTCAAAAGGCTTTCCGCTTACGCTTCCATGAGTATATG 2977
Qy 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaArgGln 960

D	2978	GCATTCGTGGCCCTTGGCGCAAAAGATCCCTTAAGGAGAGAAAGACGTCATGCTAGGCA	3037
Q	961	CysLeuValLysAsnIleAsnValArgArgLysLeuLysGlnHisAlaIleValSer	980
D	3038	TGTTTGGTGAATAATATATAATGTAAGCGGGAGATATCGAAGCAGCATGCGCTGTAGT	3097
Q	981	GluLysLeuLeuSerLeuLeuProGluLysValProTyrThrIleHisLeuLeuAla	1000
D	3098	GAATAATATATGCTCTCTTACACAGAGATGTGTTCATATACATTCACCTTTTGCA	3157
Q	1001	HisAspProAspTyrValLysValGlnAspIleGluLeuLysAspValLysGluCys	1020
D	3158	CATGACCCGAGATATATGTCMAAGTACAGCATTTGACCAACTTAAAGATGTTAAAGAAATG	3217
Q	1021	LeuTPRPheValLeuGluIleLeuMetAlaLysAsnGluValAsnAsnSerHisAlaPheIle	1040
D	3218	CTTTGGTTTGTCTCGAATATTTAATAGGCTTAAATGAAATTAACAGTCACGCTTTATTC	3277
Q	1041	ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys	1060
D	3278	AGAAAGATGATAGAAATATTTAAACAACAACAAGATGCCCAAGGACCAAGATGATGCAAAA	3337
Q	1061	MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer	1080
D	3338	ATGATATGAAACCTGATACACTGTGCTATCTTGCCATGAAATATCATCATCTCAAGAGT	3397
Q	1081	ThrTyrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln	1100
D	3398	ACTACATACACTTTGGATCTCTCTTAAGACCCCGGTACTACCAAGCTCGTTTCTTACCTCA	3457
Q	1101	ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe	1120
D	3458	CCTGCACAAAGATTTTCAGTACACCAACCAAAATTTATCTGCTCCTCGAAATGAATTCATTTTTC	3517
Q	1121	ThrProGluLysProLysThrThrAsnValLeuGluValaValAsnLysProLeuSerSer	1140
D	3518	ACTCTGAAAAACCTTAACAACAACCAATTTCTAGAGCTGTTTAACACGCCCTTTCATCA	3577
Q	1141	AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer	1160
D	3578	GCAGCAAGCATCTCAACACCAAAATCATACGAATGCAAACTGTAGCAATGCAAGCAGC	3637
Q	1161	SerSerAsnProSerSerProGluArgIleLysGluArgLeuAspSerSerGluMetAsp	1180
D	3638	AGCTCAATCCAGACTCTCTCTGAGAGATTAAGGGAGGCTTGATAGTTCGAATGGAT	3697
Q	1181	HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGluLysLysSerAspLys	1200
D	3698	CACAGTCAAAATGAAGATTACACATATGCTTCACCTTGGCCGGGAAAAAAGTGACAG	3757
Q	1201	ArgAspAspSerAspLeuValArgSerGluLeuGluLysProAlaGlyArgLysLysThr	1220
D	3758	AGAGCGACTGTGATCTTGAAGGTCTGAATTTGAGAGAGCCTAAGGACGAGAAAAAAGCG	3817
Q	1221	ProValThrGluGlnGluGluLysLeuGluIleMetAspAspLeuThrLysLeuValGlnGlu	1240
D	3818	CCCGTCACAGAACAGAGAGAAATTTAGGTATGTGATGACTTGACTAGTTGTGTACAGAA	3877
Q	1241	GlnLysProLysGlySerGlnArgSerArgLysArgLysHisThrAlaSerGluSerAsp	1260
D	3878	CAGAAACTTAAGGACGATCAGCGAAGTGCAGAAAGAGGCGCATACGGCTTCAGAAATCTGAT	3937
Q	1261	GluGlnGlnTPRPProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu	1280
D	3938	GACACGCGTGGCTCGAGGAAAGAGGCTCAAGAGAGATATATTAAGAAAAAGAAAGATCAA	3997
Q	1281	GlnAsnSerProProLysLysGlyLysArgLysArgProProLysProLeuGluGlyGly	1300
D	3998	CAGAAATATCCGCCAAAAAAGGATTAAGAAGCGCACCACAAAACCTCTGTGTGAGAGT	4057
Q	1301	ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysSerGly	1320
D	4058	ACACCAAAAGAAAGCCCAACATGATAAACTCTTAAAAAAGAGACCAAAAAAATCTGGA	4117

OY		1331	ProProAlaAProGluInGluInGluInGluInuArgInSerGlyAsnThrGluInLys	1340
Dd		4118	CCTCCAGCACACAGAGAGAGGAAGAATAAAGTGGAAATACGGAACAGAA	4177
OY		1341	SerLySserLySGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu	1366
Dd		4178	TCCAATAAGCAAAACAGCACCAGGTGTCAAGAGAGACAGCAGAGCGAATTCTCTGAA	4237
OY		1361	SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLyS	1380
Dd		4238	TCTAGTGCAATTGTAATCCACACAGTCCACACAGAAAGAGCAGGAAACATCAAAA	4297
OY		1381	ThrProSerProSerGlnProLysLySAsnVal	1391
Dd		4298	ACGCCATATCCATCATCAACAACAAAAAATGTTC	4330
RESULT 4				
ID	ABV28244			
XX	ABV28244	standard; cDNA; 7473 BP.		
AC				
XX	ABV28244;			
DT				
XX	16-SEP-2002	(first entry)		
DE				
XX		Human prostate expression marker CDNA 28235.		
KW		Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;		
KM		pharmacogenomic marker; gene; ss.		
XX				
OS		Homo sapiens.		
PN		WO200160860-A2.		
PD				
XX	23-AUG-2001.			
Pf				
XX	20-FEB-2001; 2001WO-US05171.			
PR				
XX	17-FEB-2000; 2000US-183319P.			
PR		16-MAR-2000; 2000US-189862P.		
PR		25-MAY-2000; 2000US-207454P.		
PR		09-JUN-2000; 2000US-211314P.		
PR		18-JUL-2000; 2000US-219007P.		
XX		13-DEC-2000; 2000US-255281P.		
PA				
XX		(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
PI		Schlegel R, Endege WO, Monahan JE;		
DR		WPI; 2001-662795/76.		
XX				
PT		Novel isolated nucleic acid molecule associated with cancerous state of		
PT		prostate cells and correlating with presence of prostate cancer, useful		
PT		for detecting presence of prostate cancer, stage of prostate cancer -		
PS		Claim 1; Page 5869-5870; 11750pp; English.		
XX				
CC		The invention relates to an isolated nucleic acid molecule (I) comprising		
CC		a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the		
CC		specification or its complement. (I) is useful for:		
CC		(a) assessing whether a patient is afflicted with prostate cancer;		
CC		(b) monitoring the progression of prostate cancer in a patient;		
CC		(c) assessing the efficacy of a test compound to inhibit prostate		
CC		cancer in a patient;		
CC		(d) assessing the efficacy of a therapy for inhibiting prostate cancer		
CC		in a patient;		
CC		(e) selecting a composition for inhibiting prostate cancer in a patient;		
CC		(f) assessing the prostate cell carcinogenic potential of a compound;		
CC		(g) determining whether prostate cancer has metastasized in a patient;		
CC		(h) assessing the aggressiveness or indolence of prostate cancer in a		
QC		patient;		
CC		(I) is also useful as a pharmacodynamic or pharmacogenomic marker.		
XX				

SQ Sequence 7473 BP: 2458 A: 1244 C: 1479 G: 2285 T: 7 other:

Alignment Scores:

Alignment Scores:	0	Length:	7473
Pred. No.:	7149.00	Matches:	1386
Score:	99.64%	Conservative:	0
Percent Similarity:	99.64%	Mismatches:	5
Best Local Similarity:	99.64%	Indels:	0
Query Match:	99.39%	Gaps:	0
DB:	23		

US-09-512-581b-2 (1-1391) x ABV28244 (1-7473)

QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProGlyValLys 20
DB 158 ATGGCTATTCAAAGACTAGGACCAATGATGAAAAATTCATATCCGCCGGGGTCAAG 217
QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValLys 40
DB 218 GAATATACATATAAATATCTAAAGAGAGATGCTGACACGATTAAGATGCTTGTA 277
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
DB 278 ACTTTATGGATATGACAGGACTGTGAAGAGAAAAAGAGCTTTATTAACCTAGCT 337
QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
DB 338 TTACATCTGCTTACAGATTTTCTTCACAGATCCTGATAAGATGTCCTTACGTGA 397
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
DB 398 GCCCTGCTGCTGCTGATATTTTCAGATTTATCTCTGMAAGCTCTTACACATCCCT 457
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
DB 458 GATTAACATAAGGATATATTTATCTTTATTAACAAGAGAGGAGGCTAGAGGATACA 517
QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuGluLysAsnIleAlaIleThrValLysSer 140
DB 518 AAGGCCCAATTCATACATACCTATTTTATTTACTTGAGAAACATGCTTGGGTCAAGTCA 577
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
DB 578 TATTAACATATGCTTTGAGTATGAAGATAGCAATGAATTTTACCACCAAGCTATACAGAACC 637
QY 161 LeuPheSerValIleAsnAsnGlnLysHisAsnGlnLysValHisMetHisMetValAspLeu 180
DB 638 TTATTTTCACTTATTAACAATGGCCCAATCAGAAAGTCCATATGACATGCTAGACCTT 697
QY 181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
DB 698 ATGAGCTCTATTTATTTGTGAAGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTGA 757
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
DB 758 GTAATCTGCTACCTGCTCATTAAGAAATTTAAACAAGACATATGATTTGGCAAAAGCT 817
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
DB 818 TTACTGAGAGGACAGCTCAAGCTATTTAGCCCATATTTATCCAAATTTTAAATCCAGATT 877
QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
DB 878 CTGATGCTTGGAAAAACATCTATCAGCGATTTGTCAAGACATGCTTTGACTTAAATTTTG 937
QY 261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
DB 938 GACCTCTACAAATATGATGATGCTATTTGCTCTCTCTTTTACCCCGCTGGAATTTAAA 997
QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
DB 998 TTAAAGAGCAATGATATATGAGAGCCCTTCAAGATTGTTAAACTACTGCGAAAAATGTTT 1057
QY 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuThrGlnCysTyrLeuGly 320

DB 1058 GGGCAAGGATTCAGATTCGCTTCACAAACAGCCACTTGGCAGTGTACTTGGCC 1117
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
DB 1118 AGGTTTATGATATATCCATGTCACCAATCCCGCTGGAATGTGTAATTTGGTACGCAATGCT 1177
QY 341 LeuMetAsnHisProAlaPheAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
DB 1178 CTCATGAACCACTCTGATTTTAGCAAAAGACTTAACAGAGATATCTTAAAGTAGGTCACAT 1237
QY 361 AspProGluGlnAlaIleArgHisAspValIleValSerIleValIleThrAlaAlaLysLys 380
DB 1238 GACCTCAGAGAGCTATTTAGCATGATGTTATTTGTGCANATGATACAGCTGCTTAAAAAG 1297
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
DB 1298 GATATCTCTGCTGCAATGATCATCTTAAATTTGTGAGAGAGAAACATTTAGACAAA 1357
QY 401 ArgThrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysTyrAla 420
DB 1358 CGATGAGAGAGTACGCAAAAGAGCCATGATGGAGCTTCCCAAAATTTATTAAGAAATATGCT 1417
QY 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaIleThrIleLysAspLysLeu 440
DB 1418 TTACAGTCAGAGCTGGAAGAAAGATGCTGCAAAACAGATGATGATCAAGAACAAATTG 1477
QY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
DB 1478 CTACATATATATATTCAAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1537
QY 461 GluTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
DB 1538 CAATACATAGGCTCTCCCAATTTAGAACTACACAGAGAGGAATGCTTATATTTACTTG 1597
QY 481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetCtyrLysCysGlnAsn 500
DB 1598 TATCCACACAGCTGATTTAAATGCTGTGAAGCATTTGAATGATGAATGTCAAAAT 1657
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
DB 1658 CTGCTCCGACATCAAGTAAAGATTTGCTGACTTGAATTAAGCAACCAAAACAGATGCC 1717
QY 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
DB 1718 AGTGTCAAGGCCATATTTTCAAAAGTATGCTTATTAACAAGAAATTTTACCTGATCCGTG 1777
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
DB 1778 AAGGCTCAGGATTTTCATGAAGAAATTCACACAGGCTGTAGAAATGATGAAGAAATTAAGA 1837
QY 561 LysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
DB 1838 AAGCAGTTAGCAAGTACTTGTAGTCCAAATGCTGCTGCAAGCGCTGAAGGTGTGTG 1897
QY 581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
DB 1898 CGTGAATTAACATAAGAACTTGGGCAACCCCAACAGCTTCAAAATCTTTTCTTGGAAATG 1957
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGlnSerIleSerAla 620
DB 1958 ATCAAGTTTCTCTTGGAGAGATGACACGTGTGCACATATGATACCAATCTATCAGTGCT 2017
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
DB 2018 CTTATTTAAACAAGTAAATCAATATGATGGAACACACATGATGTGAAGTGAAGGCTTT 2077
QY 641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
DB 2078 CCAACTATTAAGCCATCAAGCAGGCTTTGAACTGCTTAAAGTACTCTTATTTACAT 2137
QY 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680

Db 2138 CCCATCTCATTTTCCTGCTGAAACATTTGAAATCATTCACCTGCTGTAATAATGAT 2197
QY ASPGLuLySVaAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2198 GATGAAAAAGTAGACAGACAGCTGCATCAAAATTTTCAAAAACACAGAAACAAATTTGAA 2257
QY 701 GLuAspPheProHisIleArgSerAlaLeuProValIleuHisLysSerLys 720
Db 2258 GAGGATTTTCCACACATCAGATCAGCTTCCTGCTTTCATCATCAAAATTTAAAAA 2317
QY 721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLys 740
Db 2318 GACCCCCCGCTACACCAATATGCAATTCATGTATCATGCGATTTTCTCATGAA 2377
QY 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2378 GAGACCAGATTGTCACAGATATTGTAGCCCTGTCATTAAGAGCCTATATCCAAACCACTG 2437
QY 761 GluHisLeuIleThrProLeuValIleIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2438 GAACATCTCATTAACACCATTTGTTACTATTGGTCATATTGCTCTCTGCACTGATCAA 2497
QY PheAlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2498 TTTCGCTGCTCTTGAATCTTTGGTAGCTACTTTGATGTGAAGAATCTTCTCATGAAT 2557
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluIleLysPro 820
Db 2558 GATCGGCTTCCAGGAAAAAGACAACTTAACCTTGGGTTCCAGATGAGAAGATCTCC 2617
QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
Db 2618 GAACAAATGGTCAAAATTCAGGCTATTAAATGATGGTTCGAGGCTACTTGAATGAAA 2677
QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
Db 2678 AATTAATCACAGTAATACAGGAAGTTCTACTTAAGTTGCTAACAACAAATATGCAATG 2737
QY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2738 GATGGAGACTGTACACAAACAGGGGAAATTAAGTAAACAGATATGTGCACGCTGAGACT 2797
QY 881 AlaAlaGlySerIleAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
Db 2798 GCTGCTGGAGTCTATGTGAAGCTGGCAAGAACCCGTTTACCATGAAATCATCACA 2857
QY 901 LeuGlnGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Db 2858 TTGGAACAATATCAGCTATGTCATTAAGCATATCAAGATGAATGCTATCAAGTAAGCAA 2917
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnTyrMet 940
Db 2918 GTGTTTGCCACAGAACTTCACAAAGGCTTCCGTTTACGGCTTCCACTTGAGTATATG 2977
QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaIleArgGln 960
Db 2978 GCATATGTGCTCTTGTGCAAAAAGATCCGTAAAGAGAGAGAGACTCTGCTAGCAAA 3037
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIleAlaIleSer 980
Db 3038 TGTGTGGTGAATAATTAATGTAAAGCGGAGATATCTGGAAGCAGCATGAGCTGTATAGT 3097
QY 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3098 GAAATAATTAATGCTCTTACACAGATATGTTGTTCCATATCAATATTCACCTTTGGCA 3157
QY 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3158 CATGACCAGATATGTCAAAAGTAAGAGATATGAACAACACTTAAGATGTTAAAGAAATG 3217
QY 1021 LeuTrpPheValLeuGlnIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Db 3218 CTTTGGTTGTCTGGAATAATTAATGCTAAAAATGAATAACAGTCCAGCTTTTATC 3277

QY 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3278 AGAATAGTGTGAATAATTAATAACAAACAAAGATGCCCAAGACCAATGATGCAAAA 3337
QY 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3338 ATGAATGAATAAACTGTACACTGTGTGATGTGTCATGAATATCATGTCAAAAGACT 3397
QY 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPheThrGln 1100
Db 3398 ACTCATATACGTTTGAAATCTCTTAAGACCCGGTACTTCCAGCTGTTTCTTCACTCA 3457
QY 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluLeuMetLysSerPhe 1120
Db 3458 CCGACAAAGATTTAGTAAACACCAAAATTAATCTCTCTCGTAATGAATCAATTTTTC 3517
QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Db 3518 ACTCTGGAAACCTTAACCAACCAATGTTCAGAGCTGTAAACAAAGCCACTTTCATCA 3577
QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Db 3578 GCAGGCAAGCATCTCAGACCAATATCATCAAGATGGAACCTTAAAGCATGCAAGCAC 3637
QY 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3638 AGCTCAATTCCAAGCTCTCCTGGAAGATAAAGGAGGCTGATGATTCCTGAATAGAT 3697
QY 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3698 CACAGTGAATAATTAACAAATCTCTCACCTTCCCGGGGAAAAAGATGCAAG 3757
QY 1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
Db 3758 AGAGAGACTCTGATCTTGTAAAGTCTGAATGTGAAGAGCTTAAGAGCGGAAAAAAG 3817
QY 1221 ProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
Db 3818 CCCGTCACAGAACAGAGGAGAAATTAAGTATGATGACTGAATGGTGAACAGGAA 3877
QY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
Db 3878 CACAAACCTTAAGAGCTGACGCAAGTCGGAAGAGGCCATACGCGCTTCAGAAATCTGAT 3937
QY 1261 GluGlnGlnTrpProGluGluLysArgLeuLysGlnAspIleLeuGluLysGlnAspGlu 1280
Db 3938 GAACACAGTGGCTGAGGAAAAAGGCTCAAAAGAAATATATTAGAAATGAATGAA 3997
QY 1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
Db 3998 CAGAAATAGTCCGCCAAAAAGGCTAAAGGCGCCGACCAACCAAAACCTTGGTGGAGGT 4057
QY 1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
Db 4058 ACACCAAAAGAAAGCCAAACAAATGAATAACTTTAAAAAAAGAAACAAAAAATCTGGA 4117
QY 1321 ProProAlaProGluGluGluGluGluGluGluGlnSerGlyAsnThrGluGlnLys 1340
Db 4118 CTTCCAGCAACCAAGAGAGGAGAAAGAAAGAAAGAAAGTGAATATGCGAAGCAAG 4177
QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Db 4178 TCCAAAAGCAAAACAGACGAGTCAAGAGAGCAGACAGACAGATCTCTCGAA 4237
QY 1361 SerSerAlaIleLeuSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db 4238 TCTAGTGCATTTAAATCCACACAGTCCACACCAAAAGGAGAGAGAGACATCAAAAA 4297
QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4298 ACGGCATCACCATCAACCAAAAAAATGTC 4330

RESULT 5
ABV25469
ID ABV25469 standard; cDNA: 7473 BP.
XX
AC ABV25469;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 25460.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 5055-5057; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 7473 BP; 2457 A; 1243 C; 1481 G; 2285 T; 7 other;
SO
Alignment Scores:
Pred. No.: 0 Length: 7473
Score: 7148.00 Matches: 1386
Percent Similarity: 99.64% Conserves: 0
Best Local Similarity: 99.64% Mismatches: 5
Query Match: 99.37% Indels: 0
DB: 23 Gaps: 0
US-09-512-581b-2 (1-1391) x ABV25469 (1-7473)
QY 1 MetalHisSerLysThrArgTThrAsnAspGlyLysIleThrTyrrProGlyValLys 20
DB 158 ATGGCTCATTCAGAGCTAGACCAATGATGGAAATAATTCATATCCGCTGGGCTAAG 217
QY 21 GlutIleSerAspLysIleSerLysGluGluMetValAlaArgAlaLeuLysMetValValLys 40

DB 218 GAAATATCAGATTAATAATATCTAAAGAGAGATGGTGAACAGCATTTAAAGATGCTGTGAAA 277
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrrLeuAsnLeuAla 60
DB 278 ACTTTTATGATGATGAGACCAAGACTCTGAAAGAAAAAGAGAGCTTTATTTAAACCTAGCT 337
QY 61 LeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspValAlaArgLeuVal 80
DB 338 TTACATCTTGCTCAGATTTTCTCAAGCATCTGATTAAGATGTGCTTACTGTA 397
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrrAlaProGluAlaProTyrrSerPro 100
DB 398 GCGTGCCTCCCTTGCTGATTTTTCAGATTTATGCTCTGTAAGCTCCTTACACATCCCT 457
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGluLeuLysGlyLeuGluAspThr 120
DB 458 GATAAAGCTAAAGATATATTTATGTTTATAACAGACAGTTGAAGGGCTGAGAGATACA 517
QY 121 LysSerProGlnPheAsnArgTyrPheTyrrLeuGluAsnIleAlaIrrpValLysSer 140
DB 518 AAGAGCCACAAATTCATAGGTATTTTATTTACTTGAGAACATTCCTGGGCTCAAGTCA 577
QY 141 TyrrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrrArgThr 160
DB 578 TATAACATATGCTTTGAGATTGAAAGATAGCAATGAAATTTTACCACCACTTACAGAAC 637
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
DB 638 TTATTTTCAGTTATTAACAAATGCGCACATACAGAAAGCCATATGACATGTGAGACCTT 697
QY 181 MetSerSerIleIleCysGluGlyAspPheValSerGlnGluLeuLeuAspThrValLeu 200
DB 698 ATGAGCTCTATTTATTTGAGAGTGATACATGCTGCTGAGGCTTTTGAGATACGCTTTA 757
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrrAspLeuAlaLysAla 220
DB 758 GTAAATCTGTTACCTGCTCATTAAGATTTAAACAGACAGCATTTGATTTGGCAAGGCT 817
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrrIleThrThrPhePheAsnGlnVal 240
DB 818 TTACTGAAGAGAGACAGCTCAAGCTATGAGCCATATATTTTAAACACAGGTT 877
QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
DB 878 CTGATGCTTGGAAATCATCATCAGCATTTGCAAGCATGCTTTGACTTATTTTG 937
QY 261 GluLeuTyrrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
DB 938 GAGCTTACAAATATGATATGATGATGCTGCTGCTGCTTTTACCCGCTGAATTTAAA 997
QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
DB 998 TTTAAAGAGCAATGATATGAGAGCGCTCAAGTCTTTAAACATCACTGCGCAAAAATGTTT 1057
QY 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrrPbGlnCysTyrrLeuGly 320
DB 1058 GGGGCAAGGATTTAGATTGGCTTCTCAAAACAAGCCACTTTGGAGTGTACTGTGGGC 1117
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
DB 1118 AGGTTTATATGATATCCATGATACCAATCCGCTGGAATGTGGAATTTGGTACGACATTTGT 1177
QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrrLeuLysValArgSerHis 360
DB 1178 CTCATGAAACCATCTCGATTTAGCAAAAGACTTAAACAGATATCTTAAAGAGAGCTCAT 1237
QY 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValIrrAlaLysLys 380
DB 1238 GACCTGAGGAGGATTTAGCATGATGATTTATTTGTGCATATGTTACAGCTGCTAAAAG 1297
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValAlaArgGluAcGThrLeuAspLys 400

Db 1298 GATATTCCTTGGTCAGATGACCTTAAATTTTGTGAGAGAGAACATTAGACAAA 1357
QY ArgTPRArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
Db 1358 CGATGGAGAGTAGCGAAAAGAGCCATGATGGGACTGCCCAATTTATAGAAATATGCT 1417
QY LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTPrIleLysAspLysLeu 440
Db 1418 TTACAGTCAAGACGTGGAAAAGATGCTGCAAAAACATAGCATGATCAACAAATG 1477
QY LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
Db 1478 CTACATATATATATCAAAATAGTATGTGATGATCGACTCTGTGTAACGGATCTTGGCT 1537
QY GlnTyrMetValProHisAsnLeuGluThrThrGlnArgMetLysCysLeuTyrTyrLeu 480
Db 1538 CATACATATGGTCCCTACAAATTAGAACTACAGAAAGGATGAAATGCTTATTTACTTG 1597
QY TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTPrLysCysGlnAsn 500
Db 1598 TTATGCCACACTGGATTAAATGCTGTGAAAAGCATGTGAATGAATGTGAAATGTCAAAAT 1657
QY LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
Db 1658 CTGCTCCGACATCAAGTAAAGATTTGCTGACTGATTATTAAGCAACCCAAACAGATGCC 1717
QY SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
Db 1718 AGTGTCAAGGCCATATTTTCAAAAAGATGGTATTAACAAGAAATTTACTGTACTGCTG 1777
QY LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg 560
Db 1778 AAGGCTCAGAGATTTCATGAAGAAATTCACACAGCTGTAGGATGATGAGAAATAGCA 1837
QY LysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnAlaGluLysVal 580
Db 1838 AAGCACTTAGAAGTACTTGTGTACTCCAACTGCTCTGCAAGCAGGCTGAAGGTGTG 1897
QY ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
Db 1898 CGTGAATATACATAAGAGTTGGCCCAACCCCAACAGCTACAAATCTTCCCTGGAATG 1957
QY IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrLysIleSerAla 620
Db 1958 ATCAAACTTCTCTGTGAGAGGATAGCACCTGTGCACATAGATCCCAATCTACACTGCT 2017
QY LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluVal 640
Db 2018 CTTATTAACAAAGTGAACAAATCAATAGATGAGAACAGCATGATGATAAGAGGGGT 2077
QY ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
Db 2078 CCAACTGATCAACCCATCAGAGCAGGCTTGAAGCTGTTAAGTACTCTCATTTACACAT 2137
QY ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
Db 2138 CCCATCTCATTTCTATCTGCTGAACATTTGAATCTACTGCTGTGCAAAAAGGAT 2197
QY AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2198 GATGAAAAGGAGACAAAGCTGCATCAAAATTTTCAAAAACAGAGAACCAAAAATTGAA 2257
QY GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
Db 2258 GAGGATTTTCCACATCAGATCAGGCTTCCGTTTACATCACAAATGTAAAAA 2317
QY GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLysLys 740
Db 2318 GGACCCCGCCGTCACCAAAATATGCCATTTCATTCATTCATCGATATTTTCTAGTAAA 2377
QY GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2378 GAGACCCAGTTTGCACAGATATTGAGCCCTGCTCATAGAGCCTAGATCCAAACCAACTG 2437

QY 761 GluHisIleLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2438 GAACATCTATAACACCAATGTTGTTACTATGTGCATATGTGCTCTGCTGACCTATGCA 2497
QY PheAlaAlaProTPrLysSerTPrValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2498 TTGGCTGCTCTTGAATCTTGAATATTAATGCTTACTTTCATTTGTAAGATCTTCTCATGAT 2557
QY AspArgLeuProGlyLysLysThrThrLysIleTPrValProAspGluGluValSerPro 820
Db 2558 GATGGCTTCCAGGAAAAAGACACTTAACCTTTGGTTCCAGATGAAAGATATCTCT 2617
QY GluThrMetValLysIleGlnAlaIleLysMetMetValArgTPrLeuGluGlyMetLys 840
Db 2618 GAGACATATGTCAAATTCAGGCTATTAAGAATGATGTTGCATGCGCTACTGGAAATGAAA 2677
QY AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
Db 2678 AATTAATCACAGTAAATCAGGAACCTTCTACCTTAAGATTCCTAACAAATATATGTCATAGT 2737
QY AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2738 GATGGAGACTGACAGAACAGGGGAAATTAGTAAACCAAGATATGTACAGCTGAGACTT 2797
QY AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
Db 2798 GCTGCTGGGAGTGTCTATTTGTGAAGCTGGGCAAGAACCCCTGTTACCATGAAATCATACAA 2857
QY LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Db 2858 TTAGAACAATATCAGCTATGTGCTATGCTATCAAGATGAAGATGCTATCAAGTAAAGCAA 2917
QY ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
Db 2918 GTTGTGGCCAGAACTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAGTATAG 2977
QY AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Db 2978 GCATCTGTGGCTTGTGCAAAAAGATCCGTGAAGAGAGAAAGACCTCATGCTAGGCAA 3037
QY CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaLysSer 980
Db 3038 TGTGTGTGAAAAATTAATGAAGCGGAGATCTCAAGCAGCATCAGCTGTCTTAAT 3097
QY GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3098 GAAAAATTTATGTTCTTCTTACCAGAGTATGTTGTTCCATATACATTAACCTTTGGCA 3157
QY HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3158 CATGACCAGGATATATCAAAAGTACAGATATTAAGCAACTTAAGATGTAAAGAAATGT 3217
QY LeuTPrPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Db 3218 CTTTGGTGTGTTCTGGAAATATTAATGCGTGAAAAAGAAAATTAACAGTCACGCTTTTATC 3277
QY ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3278 AGAAAGATGTGAAAAATATTAACAAACAAAGATGCCCAAGACCAATGATGCAAAA 3337
QY MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3338 ATGAATGAAAAACCTGATACACTGTGTGATGTTGCCATGAATATCATATGTCAAAGAGAT 3397
QY ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
Db 3398 ACTACATATACGTTTGAATCTCTTAAGACCGGGTACTACAGCTGCTGTTTCTTCACTCAA 3457
QY ProAspLysAsnProPheSerAsnThrLysAsnTyrLeuProGluGluMetLysSerPhePhe 1120
Db 3458 CCTGACACAGATTTTCAGTAAACCAAAAATTAATCTGCTCTGTAATAATGAATAATCTTTC 3517

QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
 |||||
 Db 3518 ACTCCGGAAAAACCTAAACCAACCATGTTTACAGAGCTTACAGCCACTTTATATCA 3577
 |||||
 QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAsaSerSer 1160
 |||||
 Db 3578 GCAGCGAACCAATCTCAGACCAATCATCATGCAATGAAAGCTGTAAGCAATGCAACGACG 3637
 |||||
 QY 1161 SerSerAsnProSerSerProGlyArgGlyLeuLysGlyArgLeuAspSerSerGluMetAsp 1180
 |||||
 Db 3638 ACCTCAATCAACAGCTCTCTCGAAGAAATTAAGGGAGCGCTTGTATGTTCTGAAATGAT 3697
 |||||
 QY 1181 HisSerGluAsnGlnLysPylThrMetSerSerSerProLeuProGlyLysLysSerAspLys 1200
 |||||
 Db 3698 CACAGGAAATGAGATTCACATATGTTTCACTTTCCCGGGGAAAAAAGTGCACAG 3757
 |||||
 QY 1201 ArgAspAspSerSerAspLeuValArgSerGluLeuGlnLysProArgLysArgLysLysThr 1220
 |||||
 Db 3758 ACAGACGACTGTGATCTTCTTAAGGTCTGAATGGAGAACCTTAAGACGAGGAAAAAAGC 3817
 |||||
 QY 1221 ProValThrGluGlnGlnGlnLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
 |||||
 Db 3818 CCGGTCCAGAGAACAGAGAGAGAAATTAAGTATGATGACTTAAAGTTGGTACAGGA 3877
 |||||
 QY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlnSerAsp 1260
 |||||
 Db 3878 CAGAAACCTAAAGGACAGTACGAGAGTCGGAAGAACGCCATACGGCTTCAGATCTGAT 3937
 |||||
 QY 1261 GlnGlnGlnLysProGlnGlnLysArgLeuLysGlnAspIleLeuGlnLysAspGlu 1280
 |||||
 Db 3938 GACACAGCAGTGGCTGAGAGAAAAAGAGCGCTCAAGAAATATTTAGAAATGAAATGAA 3997
 |||||
 QY 1281 GlnAsnSerProProLysGlyLysAsnArgLysArgProProLysProLeuLysGlyLys 1300
 |||||
 Db 3998 CAGATATGTCGCCCAAAAAAGGCTAAAGAGCCGACCAAACTCTTGTGTGAGGT 4057
 |||||
 QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
 |||||
 Db 4058 ACACCAAAAAAGAGACCAACATGAACCTTAAAAAAAGAAAGCAAAAAAAATCTGGA 4117
 |||||
 QY 1321 ProProAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1340
 |||||
 Db 4118 CCTCCAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4177
 |||||
 QY 1341 SerLysSerLysGlnHisArgValSerArgArgIleGlnGlnArgIleGlnSerProGlu 1360
 |||||
 Db 4178 TCCAAAAGCAAAACACACGAGCTGTCAAGAGAGACACAGAGAGAGAGAGAGAGAGAGAG 4237
 |||||
 QY 1361 SerSerAlaIleGlnSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
 |||||
 Db 4238 TCTAGTGCATATGATTCACACAGTCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4297
 |||||
 QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
 |||||
 Db 4298 ACGGCATCATCATCACAACCAAAAAAATATGTC 4330
 |||||
 RESULT 6
 ABX14056
 ID ABX14056 standard; cDNA; 6744 BP.
 XX
 AC ABX14056;
 XX
 DT 24-FEB-2003 (first entry)
 XX
 DE cDNA encoding human SCC-112 tumour suppressor gene.
 XX
 ss: gene; human; chromosome 4p14; SCC-112; cancer; apoptosis; alldynia;
 KW degenerative disorder; metastasis inhibition; breast cancer; causalgia;
 KW kidney cancer; bladder cancer; pancreatic cancer; colon cancer;
 KW squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;
 KW global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia;
 KW hypoxia induced nerve cell damage; anxiety; diabetes mellitus;
 KW cardiac arrest; spinal cord lesion; stomach cancer; lung cancer;

KW neonatal distress; Alzheimer's disease; uterine cancer; colon cancer;
 KW multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;
 KW Huntington's disease; Parkinson's disease; Korsakoff's syndrome;
 KW amyotrophic lateral sclerosis; cell survival; cell proliferation;
 KW tumour suppressor.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 232..4125
 FT /tag= a
 FT /product= "Human SCC-112"
 FT polyA_signal 6721..6726
 FT /tag= b
 FT /standard_name= "PolyA signal"
 XX
 PN WO200281641-A2.
 XX
 PD 17-OCT-2002.
 XX
 XX 08-APR-2002; 2002MO-US10850.
 PF
 XX 06-APR-2001; 2001US-281780P.
 PR
 XX (GEOU) UNIV GEORGETOWN.
 PA
 XX Kasid UN, Kumar D, Ahmad I;
 PI WPI; 2003-10330/09.
 DR P-PSDB; ABG72803.
 XX
 PT New isolated SSC (undefined) tumor suppressor polypeptides and
 PT polynucleotides, useful for diagnosing, preventing or treating cancer
 PT or degenerative disease, e.g. Alzheimer's Disease, Huntington's
 PT disease, or multiple sclerosis
 PT
 PS Claim 1; Fig 1B; 83pp; English.
 XX
 CC The invention relates to a new isolated polypeptide SCC-112. The SCC-112
 CC polypeptides and polynucleotides are useful for diagnosing, preventing or
 CC treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
 CC pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
 CC cancer, stomach cancer, colon cancer, lung cancer); or degenerative
 CC disease or disorder (e.g. global and focal ischaemic and haemorrhagic
 CC stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
 CC damage, nerve cell damage caused by cardiac arrest or neonatal distress,
 CC epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
 CC pain, causalgia, neuralgias, herpes zoster, spinal cord lesions,
 CC hyperalgesia, alldynia, Alzheimer's disease, Huntington's disease,
 CC Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
 CC Korsakoff's syndrome). The polypeptides and polynucleotides are also
 CC useful for inducing apoptosis in cancer cells, increasing survival or
 CC proliferation of a cell, or inhibiting cancer cell proliferation and/or
 CC metastasis in a cancer patient. The polynucleotides can be used as probes
 CC to detect complementary nucleotide sequences, or as primers to obtain
 CC additional copies of the polynucleotides. SCC-112 may also be used for
 CC identifying drugs for treatment of cancers. The present sequence
 CC represents cDNA encoding the human SCC-112 tumour suppressor gene which
 CC is located on chromosome 4p14.
 CC
 SQ Sequence 6744 BP; 2156 A; 1243 C; 1400 G; 1945 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 0
 Score: 4760.50 Length: 6744
 Percent Similarity: 80.29% Matches: 935
 Best Local Similarity: 67.75% Mismatches: 173
 Query Match: 66.18% Indels: 210
 DB: 25 Gaps: 15
 US-09-512-581B-2 (1-1391) x ABX14056 (1-6744)
 QY 7 ArgThrAsnAspGlyLysIleThrTyrrProProGlyValLysGlnLysSerAspLysIle 26

Db 160 CGCAGACACGCCCGGAAGATCGCTTACCTCCGGGGGTAAGAGATCACGACAAATC 219
Qy SerLysGluGluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAsp 46
Db 220 ACCACGACACGAGATGATCAACGCCCTGACATGTGTGTAAGAACCTTTATGATATGAT 279
Qy GluAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
Db 280 CAGACTCAGAAAGATGAAAAACACGATATCTCCACTAGCCTTGACATCTTGATCTGAA 339
Qy PheLeuLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAsp 86
Db 340 TTCTTCCTCAGGAACCCCAATTAAGATGTGGCTCTCTTGATGATGTGTTGGCGAT 399
Qy IlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLysLeuAlaAsp 106
Db 400 ATCTTGTGATCTATGCCCCAGAAAGCTCCATATCTTCCCATGTAACCTTAAGACATA 459
Qy PheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
Db 460 TTTTGTGTTATTACACAAATTAAGGTTTGAGGATACAAAGAGCTCCACATTTAAT 519
Qy ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGlu 146
Db 520 AGATACTTTATTTATTAGAAATTTAGCTTGGGTTAAATCATATACATCTGCTTGAA 579
Qy LeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrIlePheSerValIleAsn 166
Db 580 TTGGAAATTCGAATGAATTTTATTCAGCTTTTGAACCTCTCTCTCAGTATCAAC 639
Qy AsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerIleLeuCys 186
Db 640 AATGAGCCACATTAAGAGATCAACAATGACATGCTAGATTGATGATTCATCATCATG 699
Qy GlnGlyAspThrValSerGlnGluLeuAspThrValLeuValAsnLeuValProAla 206
Db 700 GAAGGTATGAGATCTCAAGAAATTTATGGGCTCCATCTTATTAACCTCATTCCTGCA 759
Qy HisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLysArgTrpAla 226
Db 760 CATAGAACTTAATTAATAACAGTCTTGACCTTGCAAAAGTGCATTTGAAGAAGACATC 819
Qy GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThr 246
Db 820 CAGACTATGAGGATCCATTTGCAATTTTTCATCAATGATCCCTGGCTGGGAATCA 879
Qy SerIleSerAspLeuSerGlnHisValPheAspLeuIleLeuGluLeuTyrAsnIleAsp 266
Db 880 TCAGTATGATTTGTGAGAACATGTATTTGATCTGATTCAGGAACCTTTTGTATAGAT 939
Qy SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsn 286
Db 940 CCGCATTTATTAATATCCGTCATGCCACAGCTGAATTAACAATTAAGAGAATGATGGA 999
Qy GlnGluArgLeuGlnValLysLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
Db 1000 GAAGACGAGATTAAGCTGTGTTGCACTTCTAGCTTAATTTGTTGGCTCCAAAGATTCGAT 1059
Qy LeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGluArgPheAsnAspIleHis 326
Db 1060 TTGGCAACACAGAAATCGTCTCTTTGGCAATGTTTCTTGACATTTTAATGATATTCAT 1119
Qy ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
Db 1120 GTTCCCTGTAGATTAGAAAGTGTAATTTGGCCAGTATGTTTAATGAATCACCCAGAT 1179
Qy LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGlnAlaIle 366
Db 1180 TTAGCGAGAGATCTCACAGATATTTAAAGTTTGAATCACATGATCCAGAAGACTATTT 1239
Qy ArgHisAspValIleValSerIleValThrAlaLysLysAspIleLeuLeuValAsn 386

Db 1240 CGTCATGATGTCATTTGTTACTATATATAACAGCTGCCAAGAGGACCTGGCCTTAGTAAT 1299
Qy AspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLys 406
Db 1300 GATCAGCTGCTGCTTGGCTTTGTAAGGAAAGAAACATGATGATTAACCGTGGCCAGTAAGAAA 1359
Qy GlnAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGly 426
Db 1360 GAACGTATGATGGGTGTGGCTGACACTTATTAAGAAATATGCTTCTCATGTGTGAAGCAGGA 1419
Qy LysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGln 446
Db 1420 AAGGAAGCTCAGAGAAAGTCAGCTGATTAAGACCAACCTTCGATATTTATTATCAG 1479
Qy AsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis 466
Db 1480 AACGACTTGACGCAACACGCTTGATGAGAAATCTTGTCTGATCTTGTGCCCCAC 1539
Qy AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
Db 1540 AACCTGGAACAGAAAGAGAGATGAATGCTTATATTAATTAATGCTAGTTGATGATCA 1599
Qy AsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuAlaGlnVal 506
Db 1600 AATGCTTAAAGCTCTCAACGAAATGTGAAGTGAACATGCTTCGAGGCCATGA 1659
Qy LysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValAlaIlePhe 526
Db 1660 CGCAGACTATTTGATTTGGACAAAGCACCATACACAGGCTAATCTTGTCCATCTTT 1719
Qy SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMet 546
Db 1720 GGAAGACTGATGACATAGCAAGAAATTTGCTGACCCCGGAAAGCACAAGATTTTGTG 1779
Qy LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
Db 1780 AAGAAATTTAACCGGTTCTGCGCATGATGAGAACTTGGTCTGAGTTGGAGTATTA 1839
Qy ValSerProThrCysSerCysLysGlnAlaGluIleCysValArgGluIleThrLysLys 586
Db 1840 AATGAGCCCAACCTGTTCTTGCAAAACCAAGCAGATATTGTGTGAGAGAAATAGCCGGAAA 1899
Qy LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuGlu 606
Db 1900 CTTGCAAACTCTTAAGCAACCAACAATCTTTCTACAGATGCGCAAAATTTCTGTGAA 1959
Qy ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
Db 1960 AGAATCGCACCTGTGCAATGATTCAGAAAGCCCTAAGTGCACATAGGAATTTGATATAT 2019
Qy LysSerIleAspGlyThrAlaAspAspGluAspGluIleValProThrAspGlnAlaIle 646
Db 2020 AAGTCATTAAGAGGGGACAGCAGATGATGAAGAGAGGGGTAAAGTCCAGATACAGTATC 2079
Qy ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
Db 2080 CGTTCAGAGACTTGAACTTTAAGCTTCTGCTTTTACATACCTTACCTGTTGCCACTCT 2139
Qy AlaGlnThrPheGluSerLeuLeuAlaCysIleLysLysMetAspAspGluLysValAlaGlu 686
Db 2140 GCAGAGCATATGATGCTCTTGTATACAGTGGCTTAAGATGAGAGATGACAAAGTATGCGAAA 2199
Qy AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
Db 2200 GCTCTATTCAAATTTTAAAGATACAGGTGCACAAATTAAGAACAGACTTCGCCAGATA 2259
Qy ArgSerAlaLeuLeuProValIleHisHisLysSerLysGlyLysProProAlaGlnAla 726
Db 2260 CGATCGACCTTAATTTCCATTTTAACATCAAAAAGCAAGAGGGGTACTCCACACCAAGCA 2319
Qy TyrTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
Db 2320 AAACAGGCTGTGCATCTATACAGCCCATATTTCACAAATTAAGAGATTCAGCTTGCACAG 2379

QY 747 ILePheGluProLeuHISLysSerLeuAspProSerAsnLeuGluHISLeuIleThrPro 766
 DB 2380 ATTTTGAGCCACTAGTAGAGGTGTAATGCTGTATGCGCCAGAACTTATATACCTCA 2439
 QY 767 LeuValThrIleGlyHISLeuLeuLeuAlaProAspGlnPheAlaAlaProThrLys 786
 DB 2440 TTAGCTTATGGGCACATTTCTATGTTAGACAGATTCAGTTGCTTCCCAATGAA 2499
 QY 787 SerThrValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
 DB 2500 TCTGTAGTACGAATTTTATGTTGTAAGATCTGTAAATGATGACAGCTCAACAGTGA 2559
 QY 807 LysThrThrLysLeuThrPValProAspGluGluValSerProGluThrMetValLysIle 826
 DB 2560 AAGATGGAACACTGTGCTCCAGATGACAGGTTCCCTCGAAGTACTACCAAGGTA 2619
 QY 827 GlnAlaIleLysMetLeuValArgThrLeuLeuGlyMetLysAsnAsnHISerLysSer 846
 DB 2620 CAGCAATTAACCTTCTGTAAAGTGGCTGTGGTATGAAAAAACACAGCTTAATCT 2679
 QY 847 GlyThrSerThrIleuArgLeuLeuThrThrIleLeuHISerAspGlyAspLeuThrGlu 866
 DB 2680 GCCAATTAACCTTGGTTATATATATATATATATATATATATATATATATATATAT 2739
 QY 867 GlnGlyLysIleSerLysProAspMetSerArgLeuAlaArgLeuAlaAlaGlySerAlaIle 886
 DB 2740 CAAGAAGAGATCAGTAATCTGATATGTCGCTTGCATTAAGCTGCGTATGCTCCATA 2799
 QY 887 ValLysLeuAlaGlnGluProCysThrHisGluIleIleThrLeuGluGlnThrGlnLeu 906
 DB 2800 ATGAAGCTTGTCTGAGAACCTTGTATACATGAAATTTATACCCAGACAGTTTCAGCTC 2859
 QY 907 CysAlaLeuAlaIleAsnAspGluCysThrGlnValArgGlnValPheAlaGlnLysLeu 926
 DB 2860 TGTGACTTGTATTATATGATGAGTGTATCCAGTAAGGACAGATTTTGTCTCAGAGCTG 2919
 QY 927 HisLysGlyLeuSerArgLeuArgLeuProLeuGluIleuArgLeuAlaIleCysAlaLeuCys 946
 DB 2920 CATTAAGGACCTTGTGAATCTGCTCCATTTGAGATATATGCGATCTTTGCTGTGTGT 2979
 QY 947 AlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
 DB 2980 GCCAAAGATCTCTGTAAGGAGAGAACACACACACACACATGTTACTGAAAAATATC 3039
 QY 967 AsnValArgArgGluThrLeuLysGlnHisAlaAlaValSerGluLysLeuSerLeu 986
 DB 3040 AGTATACGACGAGATATACATTAAAGCAGATCTTATGCTACTGAGAAATTTATATACCTG 3099
 QY 987 LeuProGluThrValAlaProThrThrIleHisLeuLeuAlaHisAspProAspThrVal 1006
 DB 3100 TTGCTGTAATATGTAATGCTTACATGATTCACCTGCTAGCCCATGATCCCATTTTCA 3159
 QY 1007 LysValGlnAspIleGlnGluLeuLysAspValLysGluCysLeuThrPheValLeuGlu 1026
 DB 3160 AGATCACACAGATGTTGATCAGCTTCTGATATCAAGAAGTGGCTTATGCTTCACTGCTGA 3219
 QY 1027 IleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046
 DB 3220 GTTTTAATGACAAAGAAATGAAGAAACATAGCCATGCTTTAAGAAATGAGCAGCAAGAC 3279
 QY 1047 IleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuThr 1066
 DB 3280 ATCAAGATTAAACAGATGATCCCATGCTCCAGATGATCAAGACAAATGAAGAAATGTAAT 3339
 QY 1067 ThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrThrSerLeuGlu 1086
 DB 3340 ACAATATGATGATGGCTCTCTGTCTATTAATATGAAGAGCTTTGTCATCAATCCAGAT 3399
 QY 1087 SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSer 1106
 DB 3400 TCACCAAAAGACCAAGTCTCTCCCAATGAATTTTATACACAACTGAAAGAGACTTCTGT 3459

QY 1107 AsnThrLysAsnThrLeuProGluMetLysSerPhePheThrProGlyLysProLys 1126
 DB 3460 AACGATTAAGATGATATATTTCCAGAGACAGACAGACTTCTGTTAACAGGAAAGCAAG 3519
 QY 1127 ThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGln 1146
 DB 3520 CCTGCTGAGTACTAGTGTGACGATTAATTAAGCTTATATACGAAACGGAAGAAACCTAT 3579
 QY 1147 ThrLysSerSerArgMetGluThrValSer-----AsnAlaSerSerSerSerAsnPro 1164
 DB 3580 GTTAAGAACCTGCGACCTGACAGCTGGAACATATTAATGTAATTTGACGCTGAACCT 3639
 QY 1165 SerSerProGlyArgIleLysGlyArg-----LeuAspSerSerGluMetAspHisSerLys 1183
 DB 3640 TCAACCGGAATTCGATCAAGGACAGACTTCAGAGCAGACGAGAAACTGAGATTGTGA 3699
 QY 1184 AsnGluAsp-----TyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg 1201
 DB 3700 AATGAAGAGAACCTGTGAGATTAATTTACAGTACACCTGTAAGAAATATTGAC----- 3753
 QY 1202 AspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrPro 1221
 DB 3754 -----CCA 3756
 QY 1222 ValThrGlnGlnGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlnGln 1241
 DB 3757 GTAAGATTAAGGA-----ATTAAATCTGATCAAGGCTAC-----CAG 3795
 QY 1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu 1261
 DB 3796 GGCACACATCAGCAGTGCAGGAGAAAGAAAGAACAGTAAACAGCAGCTGTGCACAGAT 3855
 QY 1262 ---GlnGlnThrProGlnGlnLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
 DB 3856 ATCCACAAAAAACAGATGAGAA-----GTACATGAA 3888
 QY 1281 GlnAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeuGly--- 1298
 DB 3889 TCGGACCTCCCGCCCTCCCAACCCAGAGAGAGCTGCACCAACTCTGAATCTCAG 3948
 QY 1299 GlyGlyThrProLysGluGluProThrMetLysThrSerLysGlySerLysLysLys 1318
 DB 3949 GGCATGCTTACCAAAATGATGATCTAAATTAACCTTATTAACAGGAGGAAGAGAGACT 4008
 QY 1319 Ser---GlyProProAlaProGluGluGlu----- 1328
 DB 4009 GCAGTGGTCAAGAGAGCCGTGGGGCTTTGGAACGAGTAATGCCAAACCAACCAACTG 4068
 QY 1329 GlnGluGluArgGlnSerGlyAsnThrGlnGlnLys-----SerLysSerLys 1344
 DB 4069 CAAGATT-AGCCAAAAAGCAGACGACGAGAAAGAAATGACTTACCAAGGTAAAA 4127
 QY 1345 GlnHis-----ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSer 1362
 DB 4128 ATGCAATTTGCAAAAGGAGAAATGAAGCCCAACAGAG---CAGGCTCAGCTTGTCA 4184
 QY 4184

RESULT 7
 ABO99271 standard; cDNA; 5469 BP.
 ABO99271;
 AC ABO99271;
 XX 25-FEB-2003 (first entry)
 XX
 DE Human coding sequence SEQ ID 4.
 Human: expressed sequence tag; EST; chromosome 4;
 KW haematopoietic disorder; central nervous system disease; viral infection;
 KW peripheral nervous system disease; non-healing wound; infectious disease;
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
 KW fungal infection; autoimmune disorder; coagulation disorder; noctropic;
 KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
 KW cytosolic; haemostatic; virucide; antibacterial; fungicide;

KW Immunostimulant; cerebroprotective; gene therapy; gene; ss.
XX Homo sapiens.
OS
XX NO200259260-A2.
PN
XX 01-AUG-2002.
PD
XX
XX 16-NOV-2001; 2001WO-US42950.
XX
XX 17-NOV-2000; 2000US-0714936.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI; 2002-590824/63.
DR N-PSDB; ABP64685.
XX
XX
XX New isolated polynucleotide, useful in research, diagnostic or
PT therapeutic methods, e.g. preventing or treating disorders involving
PT aberrant protein expression or biological activity -
XX
XX
XX Claim 1; SEQ ID 4; 394pp; English.
XX
XX The present invention relates to novel human coding sequences
CC (AB099268-AB099608) and proteins (ABP64682-ABP65022). The sequences are
CC useful in therapeutic, diagnostic and research methods. The
CC polynucleotides may be used in the field of molecular biology as
CC hybridisation probes, primers for PCR, for chromosome and gene mapping,
CC for the recombinant production of protein, or in generation of anti-sense
CC DNA or RNA. The polynucleotides are useful in diagnostics as expressed
CC sequence tags (ESTs) for identifying expressed genes or for physical
CC mapping of the human genome. The proteins may be used as molecular weight
CC markers, or as nutritional sources or supplements. The proteins may be
CC used to maintain and expand cell population in a totipotentia or
CC pluripotentia state useful for re-engineering damaged or diseased
CC tissues, transplantation, manufacture of bio-pharmaceuticals or the
CC development of bio-sensors. The polynucleotides and proteins are useful
CC for preventing, treating or ameliorating disorders involving aberrant
CC protein expression or biological activity, e.g. hematopoietic disorders,
CC central/peripheral nervous system diseases, mechanical and traumatic
CC disorders, non-healing wounds, immune deficiencies and disorders,
CC infectious diseases caused by viral, bacterial or fungal infection,
CC autoimmune disorders, allergic reactions and conditions, coagulation
CC disorders, or cancer. The polynucleotide sequences of the invention were
CC assembled from ESTs isolated mainly by sequencing by hybridisation, and
CC in some cases, sequences obtained from one or more public databases.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
XX Sequence 5469 BP; 1801 A; 983 C; 1097 G; 1588 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 2.65e-202 Length: 5469
Score: 3047.00 Matches: 614
Percent Similarity: 75.53% Conservative: 133
Best Local Similarity: 62.08% Mismatches: 179
Query Match: 42.36% Indels: 64
DB: 24 Gaps: 16

US-09-512-581B-2 (1-1391) x AB099271 (1-5469)
QY 399 Asplys- --ArGtrPArGValArGlySGlAlaMeMetGlyLeuAlaGlnIleTyrLys 417
Db 42 GATRAACTTAAAGTGGCCAGTAAGAAAGAAAGCTATGATGGCTCGCTTAAAG 101
QY 418 LysTyrAlaLeuGlnInsAlaAlaGlyLysAspAlaLysGlnIleAlaTrrPLeLys 437
Db 102 AAATACGTCTTCATGCTGAACGACGAAAGAGAGCTGCACAGAAAGACACTGGATTAAG 161

QY 438 AsplysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuValGluArg 457
Db 162 GACAAACTTCGATATTTATTCACAGACAGATTACGACAAACGTTGGTAGAAG 221
QY 458 IlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeu 477
Db 222 ATCTTGTCTCAGTATCTTGTGCCACCAACCTGGAAACAGAAAGAGAAATGCGTTA 281
QY 478 TyrTyrLeuTyrAlaThrIleAspLeuAsnAlaValLysAlaLeuGlnGlnMetTrrPlys 497
Db 282 TATTACTTATATGCTAGTTGGATCCCAATAGCTGTAAAGCTCTCAACGAAATGTGGAG 341
QY 498 CysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLys 517
Db 342 TGTCAGAAACATGCTTCGAGCCATGTACGGGAACCTATTTGGATTTGGACAGACCTACA 401
QY 518 ThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPro 537
Db 402 TCAGAGCTAACGTTCTGCCATGTTTGGAAACCTGATGACATAGCAAAAGAAATTTGCTT 461
QY 538 AspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGlu 557
Db 462 GACCCCGGAAAGCACAAGATTTTGTAAAGAAATTTACCAAGGTTCTCGCGCATGATGAG 521
QY 558 LysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGlu 577
Db 522 AAATCTGCTCAGTGGATGGATTATTAATAGCCCAACCTGTTCTTCAAAACAGCAGAT 581
QY 578 GlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPhe 597
Db 582 ATTTGTGTGAGAGAAATAGCCCGAAACCTTGCAATCTTAAGCAACCAAAATCTTTT 641
QY 598 LeuGluMetLysPheLeuLeuGluArgIleAlaProValHisIleAspTrnGluSer 617
Db 642 CTGAGAGTGTCAAAATTTCTGTGGAAAGATGCAACCTGTGACATGTTATGAGAAGCC 701
QY 618 IleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAsp 637
Db 702 ATTAAGTCCCTAGCAAAATGAGCAATTAATAGCAATAGCGGACAGATGATGAAAG 761
QY 638 GlnGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuValLeuSer 657
Db 762 GAGGGTGTAGTCCAGATACAGTATCCGTTACGAGCTTACCTTAAAGTGTCTCT 821
QY 658 PheThrHisProIleSerPheHisSerAlaGlnThrPheGlnSerLeuLeuAlaCysLeu 677
Db 822 TTACACATCTCACTACCTGTTCCACTGCGAGACATATGACCTTGTACAGTGCCTA 881
QY 678 LysMetAspAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnTrnGlySer 697
Db 882 AGAATGGAGATGACAAAGTAGAGCAAGAGCTGATTCAAATTTTGAATATACAGGTCC 941
QY 698 LysIleGlnLysAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysLys 717
Db 942 AAATATGAAACACACACTTCCCCAGATPACGATCCACTTAATTCCTCATTTTACATCAAAA 1001
QY 718 SerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
Db 1002 GCAAAAGGGGTACTCCACACCAAGCAAGCAAGAGCTGTGACAGTATACAGCCATATTC 1061
QY 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
Db 1062 ACAATTAAGATTCACGCTTGCACAGATTTTGGAGCACTGAGGATGAGATGCAATGCT 1121
QY 758 SerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuAla 777
Db 1122 GATGTGCCACAACAACCTTATTAACCTCTTAGTTTACATGGCCACATTTTATGATTAAGCA 1181
QY 778 ProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeu 797
Db 1182 CCAGATCAGTTGCTTCCCAATGAATATCTGTAGTACAAATTTTATTTGGAAGAATCTG 1241
QY 798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGlu 817

```

Db      1242 CTAATGATGACAGTCAACAGTGAAAGAAAGTGAAGAACTGTGCTCCAGATGACAG 1301
Qy      818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 837
Db      1302 GTTTCCTCGAAGTACTGCAAGAGTACAGCAATTAACCTTCGTAGAGGCTGTG 1361
Qy      838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrIle 857
Db      1362 GGTATGAAACCAACCACTTAATCTGCAATTCACCTTCGGTTTATACACGATG 1421
Qy      858 LeuHisSerAspGlyAspLeuThrGlnGlnLysIleSerLysProAspMetSerArg 877
Db      1422 TTGGTAGTACGAGGAGTACAGCAAGCAAAAGAGATCAGTAATTCGATATGCTCGC 1481
Qy      878 LeuArgLeuAlaIleGlySerAlaIleValLysLeuAlaGlnLysProCysTrpIle 897
Db      1482 TTGGGATTAAGTCTGCTGAGTGCATTAATGAAGCTTGCACAGAACCTTGTACCATGA 1541
Qy      898 IleIleThrLeuGlnGlnTrpGlnLeuGlnLysAlaIleAsnAspGlnCysTrpGln 917
Db      1542 ATTATTAACCCAGAACAGTTCACAGCTCTGTGCACCTGTATTATGATGAGTGTACCA 1601
Qy      918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
Db      1602 GTAAGGAGCATATTTGCTCAGACAGCTGATAGGCATCTGTGAAGTACTGCTCCATTTG 1661
Qy      938 GluTrpMetAlaIleCysAlaLeuGlnLysAlaLysAspProValLysGlnArgAlaHis 957
Db      1662 GAGATATAGGAGCATCTTGTCTGTGCTCAAGATCTCTGTGAAGAGAGAAAGACACAC 1721
Qy      958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTrpLeuGlnHisAla 977
Db      1722 GCACGACAACTTTACTGAAATAATATCATACAGAGAGATATTAAGAGATCT 1781
Qy      978 AlaValSerGluLysLeuSerLeuSerLeuProGluTrpValValProTrpThrIleHis 997
Db      1782 ATGGCTACTGAAATATTATTATCACTGTGCTGATATGATGATCCATACATGATTCAC 1841
Qy      998 LeuLeuAlaHisAspProAspTrpValLysValGlnAspIleGlnLysLeuLysAspVal 1017
Db      1842 CTGTAGCCCATGATCCAGATTTTACAGATCACAGATGTTGATCAGCTTCGTATATC 1901
Qy      1018 LysGluCysLeuTrpPheValLeuGlnIleLeuMetAlaLysAsnGlnLysAsnSerHis 1037
Db      1902 AAAAGATGCTATGCTTCATCTGATGATTTAATGACAAAGAAAGAAACAAATGCCAT 1961
Qy      1038 AlaPheIleArgLysMetValGlnAsnIleLysGlnThrLysAspAlaGlnLysProAsp 1057
Db      1962 GCCTTTTGAAGAGATGAGGAGAGACATCAAGTTAACACAGATGCCACAGCTCCACAT 2021
Qy      1058 AspAlaLysMetAsnGlnLysLeuTrpValLysAspValAlaMetAsnIleIleMet 1077
Db      2022 GAATCCCAAGACAAATGAAACCTGATACATATGATGATGCTCTGTGTATTAAT 2081
Qy      1078 SerLysSerThrTrpTrpSerLeuGlnSerProLysAspProValLeuProAlaArgPhe 1097
Db      2082 AGTAAAGTCTTTGTGCAATGCAGATTCACCAAGAGCCCAAGCTCTCCCAATGAAATTT 2141
Qy      1098 PheThrGlnProAspLysAsnPheSerAsnThrLysAsnTrpLeuProGluMetLys 1117
Db      2142 TTTCACACACCTGAAAGAGACTCTGTACATGAATGAATATTTCAGAGAGACAGCA 2201
Qy      1118 SerPhePheThrProGlyLysProLysTrpThrAsnValLeuGlyAlaValAsnLysPro 1137
Db      2202 GTACTTCTGTTAACAGAAACCAAGCCCTGCTGAGACTAGGCTGCTAAATTAAGCCT 2261
Qy      1138 LeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
Db      2262 TTATCAGCAACGGAAGAAACCTATATGTTAGAAACACTGCACACTGAACTGGAAGCAAT 2321
Qy      1157 ---AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg---Leu 1174

```

```

Db      2322 ATTAATGTAATTCAGACCTGAACCTTCAACCGGAAATCGATCAAGGAACAGATTCA 2381
Qy      1175 AspSerSerGluMetAspHisSerGlnGlnLysAsp-----TyrThrMetSerSerPro 1192
Db      2382 GAGGACGACGAAATCGAGTGTAGTGAATAATGAAGAAACCTGTGAGGATTAATTCACATC 2441
Qy      1193 LeuProGlyLysLysSerAspLysArgAspSerAspLeuValArgSerGluLeuGln 1212
Db      2442 ACACCTGTAAAGAAATATGAC----- 2462
Qy      1213 LysProArgGlyArgLysLysThrProValThrGlnGlnGlnLysLeuGlnLysMetAsp 1232
Db      2463 -----CCAGTAAAGAAATGAAGAA-----ATTAATTCGTAT 2492
Qy      1233 AspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArgLysArg 1252
Db      2493 CAGGCTTACC-----CAGGCAACATCAGCAGTACAGCCAGAGAAAGAAAGA 2537
Qy      1253 GlyHisThrAlaSerGlnSerAspGlu---GlnGlnTrpProGluGlnLysArgLeuLys 1271
Db      2538 ACAGTACACGACGCTGTGTGACAGATATCCACAAACAAACAGATGAGAAA----- 2588
Qy      1272 GluAspIleLeuGlnLysAsnGlnLysAsnSerPro---ProLysLysGlyLysArg 1290
Db      2589 -----GTAGATGAATCGGAGACTCCGCCCTTCCAAACCCAGAGAGA 2630
Qy      1291 GlyArgProProLysProLeuGly---GlyGlyThrProLysGlnGlnLysProThrMetLys 1309
Db      2631 GCACCTGCACCCCACTGAAATCTCAGGCAATGCTACCAAAATGAATGATCTTAATTA 2690
Qy      1310 ThrSerLysLysGlySerLysLysSer---GlyTrpProAlaProGluGlnLysGln 1328
Db      2691 CCTATTACAAAGGAGAGAGAGACCTGACAGTGGTCCAGAGAGCCCTGGGGGTTTGAA 2750
Qy      1329 -----GlnGlnLysArgGlnSerGlyAsnThrGln 1339
Db      2751 GCAGTAAATGCCAAACCAACCACTGCAAGATTT-ACCCAAAAAGGACAGCCAGCAGAGA 2809
Qy      1340 Lys-----SerLysSerLysGlnHis-----ArgValSerArgAlaGln 1353
Db      2810 AAGACAAATTAATGACTTACAAAGTAAATGATTTGCAAGAGGAGAGAAATGAAGCCAAA 2869
Qy      1354 GlnArgAlaGlnSerProGluSerSer 1362
Db      2870 CAGAG---CAGGCTCCAGCTTCTGCA 2893

RESULT 8
AAA47423 standard; DNA; 3957 BP.
AAA47423:
20-OCT-2000 (first entry)
Sequence encoding human neuron-associated protein.
Neuron associated protein; NEUP; neurological disorder; epilepsy;
ischemic cerebrovascular disease; stroke; cerebral neoplasm;
Alzheimer's disease; Pick's disease; Huntington's disease;
dementia; Parkinson's disease; demyelinating disease; meningitis;
prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
cerebral palsy; muscular dystrophy; central nervous system; CNS;
peripheral nervous system; PNS; myopathy; schizophrenia;
actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
autoimmune disease; inflammation; acquired immunodeficiency syndrome;
AIDS; Addison's disease; adult respiratory distress syndrome;
allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
Werner syndrome; trauma; human; ds.
Homo sapiens.
XX
OS

```

Key	Location/Qualifiers
FT CDS	140..2704
FT	/*tag= a
FT	/product= Neuron associated protein
XX	
PN	MO200034477-A2.
PD	15-JUN-2000.
XX	
PF	10-DEC-1999; 99MO-US30408.
XX	
PR	11-DEC-1998; 98US-0210083.
PR	11-DEC-1998; 98US-9123456.
PR	09-FEB-1999; 99US-0119365.
PR	16-MAR-1999; 99US-0124687.
XX	
PA	(INCYT-) INCYTE PHARM INC.
XX	
PI	Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
PI	Lu DM, Azimzal Y;
XX	
DR	WPI; 2000-423423/36.
DR	P-PSDB; AAB01382.
XX	
PT	New human neuron-associated proteins and polynucleotides encoding them,
PT	useful for diagnosis, treatment and prevention of cell proliferative
PT	disorders including cancer, neuronal and neurological disorders
XX	
PS	Claim 9; Page 130-131; 145pp; English.
XX	
CC	Human neuron-associated proteins (NEUP) can be used for for
CC	treating or preventing a disorder associated with decreased
CC	expression or activity of NEUP. Antagonists of NEUP are useful for
CC	treating or preventing disorder associated with increased expression
CC	or activity of NEUP. NEUP or their fragments or derivatives are
CC	useful for treating neurological disorder such as epilepsy, ischemic
CC	cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC	disease, Pick's disease, Huntington's disease, dementia and
CC	Parkinson's disease. NEUPs are also useful for treating other
CC	demyelinating diseases, bacterial and viral meningitis, prion
CC	diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC	metabolic diseases of the nervous system, neurofibromatosis, other
CC	developmental disorders of the central nervous system, cerebral
CC	palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC	cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC	other neuromuscular disorders, peripheral nervous system disorders,
CC	inherited, metabolic, endocrine, and toxic myopathies, mental
CC	disorders including mood, anxiety and schizophrenic disorders, a cell
CC	proliferative disorder such as actinic keratosis, arteriosclerosis,
CC	atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC	tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC	hemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC	bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC	disease, adult respiratory distress syndrome (ARDS), Addison's
CC	spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC	complications of cancer, hemodialysis, and extracorporeal circulation,
CC	viral, bacterial, fungal parasitic, protozoal, and helminthic
CC	infections, and trauma. This sequence was given the Incyte ID no.
CC	2888437CB1.
XX	
SQ	Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;
XX	
Alignment Scores:	
Pred. No.:	1 66e-194
Score:	2932.50
Percent Similarity:	74.51%
Best Local Similarity:	60.90%
Query Match:	40.77%
DB:	21
US-09-512-581b-2 (1-1391) x AAA47423 (1-3957)	

QY	428	ASpAlaAlaLysGlnIleAlaTrrPLeuLysAspLysLeuLeuHisIleTrrYrGlnAsn	447
Db	2	GAAGCTGACGAGAAAGACACCTGCGATTAAGAGCAAACTCTGATATTATATATCAGAAC	61
QY	448	SeTLeaspAspArgLeuLeuValGluArgIlePheAlaGlnPyrMetValProHisAsn	467
Db	62	AGCATTTGACGACAAACCTGTTGGTAGAGAAATCTTGGCTCAGTATCTGGCCCAAC	121
QY	468	LeuGlnThrThrGlnMetLysCysLeuTrrYrLeuTrrAlaThrLeuAspLeuAsn	487
Db	122	CTGGAACAGCAAGAGAAATGAATGCTTATATCTTATCTGATTTGATTCGAAT	181
QY	488	AlaValLysAlaLeuAsnGluMetTrrPLeuCysGlnAsnLeuLeuArgHisGlnValLys	507
Db	182	GCTGTAAAGCTCTCAACGAAATGTGGAAGTGCACAAACCTGCTCGATCCATGACCC	241
QY	508	AspLeuLeuAspLeuLeuLysGlnProLysThrAspAlaSerValLysAlaIlePheSer	527
Db	242	GAACCTATGTGATTGGACAAAGACGCTACATCAAGAGCTAACTGTTCTGCAATTTGGA	301
QY	528	LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys	547
Db	302	AAACTGATGACCATGACCAAGAAATTTGCTGACCCGGGAACACAGATTTTGGAG	361
QY	548	LysPheThrGlnValLeuGlnAspAspGlnLysIleArgLysGlnLeuGlnValLeuVal	567
Db	362	AAATTTAACCGAGTCTCGGCGATGATGAAGAACTCGCTCAGTTGAGTTATTAAT	421
QY	568	SerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeu	587
Db	422	AGCCCAACCTGTTCTTGCACAAACAGAGATTTGTGTGAGAAATAGCCCGGAACCT	481
QY	588	GlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPheLeuLeuGluArg	607
Db	482	GCAATTCCTAAGACCAACAAATCTTTCTAGAGTGTCAATTTCTGTTGGAAGA	541
QY	608	IleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLysGlnValAsnLys	627
Db	542	ATCGCACCTGTGACATGATTCATGAGAACCATTAAGTGCACCTAGTGAATGAATTAAC	601
QY	628	SerIleAspGlyThrAlaAspAspGluAspGluGlyValAlaProThrAspGlnAlaIleArg	647
Db	602	TCAATAGAGGGGACAGACAGATGATGAAGAGAGGTTAGTGCAGATACGATACCT	661
QY	648	AlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAla	667
Db	662	TCAGAGCTTAACCTTTAAGTGTCTGCTTTTAACATCTTCACCTGCTCCACTGCA	721
QY	668	GlnThrPheGlnSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAla	687
Db	722	GAGACATATGAGCTCTTGTTCAGTGCCTTAAGATGAGATGACAAAGTAGCAGAACT	781
QY	688	AlaLeuGlnIlePheLysAsnThrGlySerLysIleGluLysPheProHisIleArg	707
Db	782	GCTATTTAAATTTTAAATAATACAGGTCAACAAATTAAGACACCTTCCCAATACGA	841
QY	708	SerAlaLeuLeuProValLeuHisLysSerLysGlyProProArgGlnAlaLys	727
Db	842	TCGACCTTAATTCCTATTTACATCAAAAAGCAAAAGGGGTACTCCACACCAAGAAA	901
QY	728	TrrAlaIleHisCysIleHisAlaIlePheSerSerLysGlnThrGlnPheAlaGlnIle	747
Db	902	CAGGCTGTCAGTATACAGCCCATATTCACAAATTAAGCAAGCTGTCAGACAGATT	961
QY	748	PheGlnProLeuHisLysSerLeuAspProSerAsnLeuGlnHisLeuIleThrProLeu	767
Db	962	TTTGAGCACTCAGTAGAGCTGGAAGCTGATGTGCCAACAACTTAATATCACTTA	1021
QY	768	ValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaIleProTrrPLeuSer	787
Db	1022	GTTTCATTTGGGACCAATTTCTATGTTAGCCACCAATCATGTTGCTTCCCAATGAATCT	1081
QY	788	TrrValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys	807


```

Db      1082  GTAGAGCAAAATTTATGTGTAAGATCTGCTAATGATGACAGGTCAACAGCTAAAG 1141
Qy      808  ThrThrLysLeuTrpValProAspGluGlnValSerProGluThrMetValLysIleGln 827
Db      1142  AATGAAAGCTGTGCTCCAGATGAGAGGTTTCCCTGAAAGTCAAGTACAG 1201
Qy      828  AlaIleLysMetMetValArgTrpLeuGluGlyMetLysAsnAsnHisSerLysSerGly 847
Db      1202  GCAATTTAACTTCGTGTAAGGCTGTGGTAAAGAAAACACACAGCTTAAATCTGCC 1261
Qy      848  ThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlnGln 867
Db      1262  AATTCACCCCTTCGGTTTATATACAGCATCTTGGTTAGAGGGTGACCTGCACAGCAA 1321
Qy      868  GlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaIleGlySerAlaIleVal 887
Db      1322  AAGAGGACAGTAAATCTGATATGTCCTGCTGCATTAAGCTGTGATGCTCCATTAATG 1381
Qy      888  LysLeuAlaGlnGluProCysTrpHisGluIleIleThrLeuGluGlnGlyTrpGlnLys 907
Db      1382  AAGCTTGCTCAGGAACCTGTGTACCATGAAATTAATACCCAGAACAGTTTCAGCTCTGT 1441
Qy      908  AlaLeuAlaIleAsnAspGluCysTrpGlnValArgGlnValPheAlaGlnLysLeuHis 927
Db      1442  GCACCTGTTATTAATGATGAGGTACCAAGTAAGCAGATATTGTCACAGAGCTGCAT 1501
Qy      928  LysGlyLeuSerArgLeuArgLeuProLeuGluIuTrpMetAlaIleCysAlaLeuGlySala 947
Db      1502  AAGGACCTTGTAAGTACTCTCCCATTTGAGTATATGGCATCTTTGCCCTGTGTGCC 1561
Qy      948  LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsn 967
Db      1562  AAGAGTCTGTGAGAGAGAGAAAGACACACGACACATGTTTACTGAAATAATTCAGT 1621
Qy      968  ValArgArgGluTrpLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeu 987
Db      1622  ATAGCGCAGGAATACATTAAACAGAAATCTGCTACAGAAATTTATACATGCTTGG 1681
Qy      988  ProGluTrpValValProTrpThrIleHisLeuLeuAlaHisAspProAspTrpValLys 1007
Db      1682  CCTGAAATGTAGTCCATACATGATTCACCTGCTAGCCCATGATGCCAGATTTTACAA 1741
Qy      1008  ValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIle 1027
Db      1742  TCACAGAGTGTGATCAGCTTCGTGATATCAAAAGGTCCTATGATGCTTATGTAAGT 1801
Qy      1028  LeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1047
Db      1802  TTAATGACAAAGAAATGAAACAAATAGCCATGCTTATGAAGAAATGCGACAGAAACATC 1861
Qy      1048  LysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGlnLysLeuTrpThr 1067
Db      1862  AAGTTAAACCAAGATGCCCTCCAGATGATCCCAAGACAAATAAGAAACTGTATACA 1921
Qy      1068  ValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTrpSerLeuGlnSer 1087
Db      1922  GTAATGTATGTGCTCTCTGTGTTATATAGTAAAGTGCCTTGTGCAATGCAATTCAC 1981
Qy      1088  ProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsn 1107
Db      1982  CCAAGAGCACCAAGCTCCCAATGAATTTTTCACAACTGAAAGAGACTTCGTATAC 2041
Qy      1108  ThrLysAsnTrpLeuProGluMetLysSerPhePheThrProGlyLysProLysThr 1127
Db      2042  GATTAAGACTTATATTTCAGAAAGACAGAGTACTTCTGTTAAACGAAAGCAAAAGCCT 2101
Qy      1128  ThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyGlnSerGlnThr 1147
Db      2102  GCTGAGTACTAGGTGCAGTAAATRAAGCTTTATACAGCAAGGAGGAAACCCATGTT 2161
Qy      1148  LysSerSerArgMetGluThrValSer-----AsnAlaSerSerSerAsnProSer 1165
:::|||||: ||||| ||| ::||| |||||

```

```

Db      2162  AGAAGCACTGGCACTGAGACGTGAAGCAATATTAAATGTAATTCAGAGCTGAACCTTCA 2221
Qy      1166  SerProGlyArgIleLysGlyArg---LeuAspSerSerGluMetAsnHisSerGluAsn 1184
Db      2222  ACCGGAATTCGATCAAGGAAAGGACGATTCAGAGGACGAGAAACTGAGATTGTGAAAT 2281
Qy      1185  GluAsp-----TyThrMetSerSerProLeuProGlyLysLysSerAspLysArgAsp 1202
Db      2282  GAAAGAAACCTGTGAGATTTATTTTCAGTCACACCTGTAAGAAATATTGAC----- 2332
Qy      1203  AspSerAspLeuValArgSerGluLeuGlnLysProArgGlyArgLysLysThrProVal 1222
Db      2333  -----CCAGTA 2338
Qy      1223  ThrGlnGlnGluGlnLysLeuGlyMetLysAspLeuThrLysLeuValGlnGlnLys 1242
Db      2339  AAGAAATAAGAA-----ATTAAATCTCATCTACAGCTACC-----CAGGCG 2377
Qy      1243  ProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu--- 1261
Db      2378  AACATCAGCAGTGACCGAGAAAGAAAGAAACAGATACAGAGCTGTGTCAGAGATATTC 2437
Qy      1262  GlnGlnTrpProGlnGluGlnLysArgLeuLysGluAspIleLeuGluAsnGluAspGln 1281
Db      2438  CAACAAAAACAGATGAGAAA-----GTAGATGAATCG 2470
Qy      1282  AsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeuGly---Gly 1299
Db      2471  GGACCTCCCGCCCTCCCAAACCCAGAGAGACCTCGACCAAGTGTGAATTCACAGGCG 2530
Qy      1300  GlyThrProLysGluGlnProThrMetLysThrSerLysGlySerLysLysSer 1319
Db      2531  AATGCTACCAAAATGATGATCTTAATTAACCTTAACAAAGGAGAGAGAGAGCTGCA 2590
Qy      1320  --GlyProProAlaProGlnGlnGlnGlu-----Glu 1329
Db      2591  GTGGGTCAAGAGAGCCCTGGGGGTTTGGACAGCAGTAATGCCAAAGCACCAACTGCAC 2650
Qy      1330  GlnGluArgGlnSerGlyAsnThrGlnGlnLys-----SerLysSerLysGln 1345
Db      2651  GATTT-AGCCAAAAAGGAGCAGCCAGCAAGAAACAAATGACTTACAAAGGTAAATAATG 2709
Qy      1346  His-----ArgValSerArgArgAlaGlnGlnArgAlaGlnLysSerProGluSerSerAla 1363
Db      2710  CATTTGCCAAAGGAGGAAATGAAGGCCCAACAGAAAG---CAGCTCCAGCTTCTCCAAA 2766
Qy      1364  -----IleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
Db      2767  ACTTGATTCACAAATGTCCCTGAACAGAAATGAAGCTCAACGACGACG 2817

RESULT 9
ABL22105
ID      ABL22105 standard; DNA; 3920 BP.
XX      XX
AC      ABL22105;
XX      XX
DT      26-MAR-2002 (first entry)
XX      XX
DE      Drosophila melanogaster genomic polynucleotide seq ID NO 17788.
XX      XX
KW      Drosophila; developmental biology; cell signalling; insecticide;
XX      XX pharmaceutical; gene; ds.
OS      Drosophila melanogaster.
XX      XX
PN      WO200171042-A2.
PD      27-SEP-2001.
XX      XX
PF      23-MAR-2001; 2001WO-US09231.
XX      XX
PR      23-MAR-2000; 2000US-191637P.
PR      11-JUL-2000; 2000US-0614150.

```


Dd	1887	ACCACGTTAGCAACAACTCGGAAACATGCTACGCTCCATACATGATTAACATTCCTC	1946
Oy	567	ValSerProThrCysSerCysAlaGlnAclGluCysValArgGluIleThrLysLys	586
Dd	1947	CTTAAACCGACAGTAACTGTCTGGGAGTGTCTGACAGAGGGAGTCTCTCGTGAAGAAG	2006
Oy	587	LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuGlu	606
Dd	2007	CTTGCGCGCCACGTCCAA---TCCAAATCTGATTACAAACAGTTAAATGCTGGATTGAG	2063
Oy	607	ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn	626
Dd	2064	CGCGTGGCATCGATGATGTGGACAAAGAGCTATTGGCGTACATTATTAAGCTTAATGA	2123
Oy	627	LysSerIleAspGlyThrAlaAspAspLysArgLysValProHisAspGlnAlaIle	646
Dd	2124	CAATTCATCTCGAAGAAAGGGTCATGTGTAAAGATCGGAATTTGGCTCAGAAAGCGGC	2183
Oy	647	ArgAlaGlyLeuGlnLeuLysValIleSerPheThrHisProIleSerPheHisSer	666
Dd	2184	GAGCGGAGACATCAAGTTGTTGGCATGCTTTCATACGCTTTTGGCGCTATTCTTCACC	2243
Oy	667	AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGlyLysValAlaGlu	686
Dd	2244	GATACCTTCGCTGCGCTAATTTAATTTCCCTGCTACGAGCAGATTAATGTGACCA	2303
Oy	687	AlaIleLeuGlnIlePheLysAsnThrGly-----SerLysIleGluGlnAspPhe	703
Dd	2304	CTGGTGCCTAAAGACACTAATCATCTGGGGCGCTATACGCCATCATTTGACGACCTTACA	2363
Oy	704	ProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysGlyProPro	723
Dd	2364	CCGCGTATTCTCGACGACACTGCGTCGGGTGTGCAAGGACTTGGCACTAATTGGAACCCC	2423
Oy	724	ArgGlnAlaLysTyraIleHisCysIleHis-----	734
Dd	2424	AAGCAAGCAAAAGCACGGGTGCGATGTATTTCTTAACACGACAGTGTGCGGCTCCACT	2483
Oy	735	-----AlaIlePheSerSerLysGluThrGln-----	743
Dd	2484	GATGGAGCAACAAGTGGAGCTGGAAAGTCCGTCACAACTACACAAGAGTGCATTCATT	2543
Oy	744	PheIleGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGlnHisLeu	763
Dd	2544	TTTCAACGAGCATCTGAGACCTTGGCTGCCCTAAATGTGACACCA---AACTGTGACATTCAG	2600
Oy	764	IleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaIle	783
Dd	2601	CGCACAAAGATTGTGACCTTGGGTGCACATTGCCATACATATGCCACAGGCTTCTCTAACG	2660
Oy	784	ProThrLysSerThrPvalAlaIleThrPheIleValLysAspLeuLeuMetLysAspArgLeu	803
Dd	2661	CCCATTTAAAGATTATGATTGATCGCCGACGAATGTCTCAAAAGAGCTGTATTCACGAAGA---GT	2717
Oy	804	ProGlyLysLysThrThrLysLeu-----ThrValProAspGluGluVal	818
Dd	2718	CTTCGCGACGACACTACGAACCTCCACAGGACAGTGATTTGGTGTGCCCAAGAAACTA	2777
Oy	819	SerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeuGlu	838
Dd	2778	CCGCGCGGACATCTATATACAACTGGATGCCCTCAAGAGGTATAGGCCAGGTGCTTTGGGA	2837
Oy	839	MetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeu	858
Dd	2838	TTTGGTACCGGAT---GAGCACGCTGCCCCAGAAACATTCGGAATGTGGCGGCTTTGT	2894
Oy	859	HisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeu	878
Dd	2895	AACCAACAGGCGATTTGCTGGCCACAGACCTCTTTGCGGTGCCGGAATAATCTTGGCTG	2954
Oy	879	ArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrlHisGluIle	898

Db	2955	CGCCCTGGGGGACGCTGGCGCATGCTCAAAGTGTGCGAAACAAAGGGCGTGAAGTATGAC	3014
Qy	889	IlIeThLeuGIuGIuThrGIuLeuCysAlaIeAlaIleAsnAspGIuCyTfGIuVal	918
Db	3015	TACAGCGGTGACACTATTTTTCAGCTTTCCAGCTGATGGGTGATCCGGTCCAGAACTT	3074
Qy	919	ArgGInVaIlePheAlaGlnIleLysIleuIstIysGIuLeuSerArgLeuArg-----	934
Db	3075	CGGGAAATCTTTGGCTGCGCAAGCTGCACAAAGGATTAAAGCAGAAGTTGCCAGGAACGT	3134
Qy	935	LeuProLeuGIuTfYrMetAlaIleCysAlaLeuCysAlaLysAspProValIysGIuArg	954
Db	3135	TTGGCGCGTGACCTTCATGGCTTTGATGTGCTGCTGGCTGTACAGACTGAGAGCAAAATTG	3194
Qy	955	ArgAlaIstAlaIstArgGlnCysLeuValIysAsnIleAsnValIstArgGIuTfYrLeuLys	974
Db	3195	CAGACGCTTGTCGGCTACTATGACAGAAAGGATGTAAACAAAGGGGGGAAATATCTCAAG	3254
Qy	975	GlnHstAlaIstAlaValSer-----GIuLysLeuLeuSerLeuLeuPro	988
Db	3255	ACTGTCGGCTATGACATCTTCCTCCGACAGCTCAACGGAGTCACATATCTACATACTACT	3314
Qy	989	GIuTfYrValIstValProTfYrThIleIstLeuLeuAlaIstAspProAspTfYrValIstAl	1008
Db	3315	GACTCAATGCTGGCTTGTGCTATTCCTGCTGGTCCAGATCCACGCTTACAGAAATCAC	3374
Qy	1009	GlnAspIleGIuGlnIleuLysAspValIysGIuCyLeuTfPheValIleuGIuIleu	1028
Db	3375	GAGGACTACGTACAGCTGCTGCCAGATGAGAGATGCTTCGCTTATCTTGAGCCGCTG	3434
Qy	1029	MetaLalYsAsnGIuAsnAsnSerIstAlaPheIleArgLysMetValGIuAsnIleLys	1048
Db	3435	ATGGCCAAACGAGAAACGCTTGTCCATGTCTTCAACGACAGCTGCTGACCTGTAAAG	3494
Qy	1049	GlnThrLysAspAlaGlnGIuProAspAlaLysMetAlaGlnLysLeuTfYrThVal	1068
Db	3495	CATCGCGAGTTTCAGTCTGGGG---TCGACAAAGCGCGACACATATAAATGTGGCGCTC	3555
Qy	1069	CysAspValAlaIstAlaMetAsnIleIleMetSerLys-----SerThr	1081
Db	3552	TGGCATCTTTCGCATGATATCATCATGACATCCCAAGTTCACATTTGATGGCAACAGAGC	3611
Qy	1082	ThrTfYrSerLeuGIuSerProLysAspProValLeuProAlaIstArgPheThGIuIstPro	1101
Db	3612	ACCTTTTCCATGCGCGCTG-----GCTTTTCCAGAAATGATTATTAAGAGCTT	3655
Qy	1102	Asp---LysAsnPheSerAsnThrLysAsnTfYrLeuProGIuMetLysSerPhePhe	1120
Db	3660	GCGCTTGGCAATTTCCAAACAAATGACGTATATACCGCTGAGACGTG-----TAT	3710
Qy	1121	ThrProGIuLysProLysThrThAsnValIleuGIuAlaValAsnLysProLeuSerSer	1140
Db	3711	ACGCTGGGAGCCAAATTCACAGAGCAACGCTGCCGACAGC-AATGACACAGCTGCCGAGC	3769
Qy	1141	AlaGIuLysGlnSerGIuThr-----LysSerSerArg	1151
Db	3770	AGCACTGGCTCCAAAGAGACCGGGCGGAACAGTCAATCATGTGATGTAATTAATTCGACAGA	3829
Qy	1152	MetGIuThrValIstAsnAlaIstSerSerSerSerAsnProSerSerProGIuArgIleLys	1171
Db	3830	GACACAACTGTGTTGCACAACTACAGAGGGGTGACACACTACGAGAGCCCATGGCAAAAGAAC	3889
Qy	1172	GlyArg	1173
Db	3890	GCGCGC	3895
RESULT 10			
ABL22104/C			
ID	ABL22104	standard: DNA: 6950 BP.	
XX	ABL22104;		
DT	26-MAR-2002	(first entry)	

XX XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 17785.
 DE XX
 XX Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmacological; gene; ds.
 XX OS Drosophila melanogaster.
 XX PN MO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001MO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PMD, Myers EW;
 XX DR WPI; 2001-656860/75.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX PS Claim 1; SEQ ID NO 17785; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 6990 BP; 1907 A; 1513 C; 1586 G; 1984 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,48e-110 Length: 6990
 Score: 1729.50 Matches: 463
 Percent Similarity: 49.43% Conservative: 270
 Best Local Similarity: 31.22% Mismatches: 436
 Query Match: 24,04% Indels: 318
 DB: 23 Gaps: 37
 US-09-512-581b-2 (1-1391) x AB122104 (1-6990)
 QY 13 IletHrThyrProFroglyValLysGluIleSerAspLysIleSerLysGluIleMetVal 32
 DB 5398 ATAGTATACCCACCCGATGCCGCGCTGCTGAGGATTTGGGACACGACGATGATA 5339
 QY 33 ArgArgLeuLysMet-ValVal----- 39
 DB 5338 AGCGCGCTCAAGGTAGTCTCGCCCTGTTCCCTTAGCAAAACTTCTCATTAGATGA 5279
 QY 40 -----LysThrPheMetAspMetAspGlnAspSerGluGlu 52
 DB 5278 TTCTGTTTACAGACCTCGCCCAATGTTCTGCAGACCATGACGACGACCAATCTCTA 5219
 QY 52 ulysGluLeuTyLeuAsnIleuHisLeuAlaSerAspPhePheLeuLysHisPr 72
 DB 5218 CCAAGCAA---TACATACCATGCGCCCTCCATCTGCTGGAGCACTTCTTATGCAACATCC 5162
 QY 72 ogLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyra 92
 DB 5161 ATCCCGTGAATGCTCAGACTCTTATTCGCGTCTCGTGGCGGATGTCTCTGCGGTTATGC 5102

QY 92 aProGluAlaProTyThrSerProAspLysAspIlePheMetPheIleThrAr 112
 DB 5101 GCCAGAGGCTCCCTTACAAAGACGACGATCAAAATCAATATTTAAGTTTATTCAA 5042
 QY 112 gGluLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyThrLeuLe 132
 DB 5041 GCAACTGCAAGGCTGTAAGGATCGCGCGATCCCTTTCAGCGTACTCTTATCTGCT 4982
 QY 132 uGluAsnIleAlaIlePheLysSerTyraAsnIleCysPheGluLeuGluAspSerAsn 152
 DB 4981 GGAGAACCTTCTCTTTTCTCAAGTCTTTATATGTCTTTGCAACTGGAGACTGCCAAG 4922
 QY 152 ullePheThrGluLeuTyraArgThrLeuPheSerValIle----- 165
 DB 4921 AATCTTCAAGACCTGTTAGTACATCTTCAGATGTCAGATGCAAGTCAGACTCTTCT 4862
 QY 166 -----AsnAsnGlyHisAsnG 171
 DB 4861 TTATCTCGCACTCTTATTAATCAGATGTATATAATTTTCTAGTACAGCAGCATGT 4802
 QY 171 nLysValHisMetHisMetValAspLeuMetSerIleIleCysGluGlyAspThrVa 191
 DB 4801 CAAGGTACTAATCTTTTCTAGATGTGCTAAGCCGTGTATCAGACGACGACATCT 4742
 QY 191 lSerGluGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
 DB 4741 GTGCGTGAATTTGTGATCTCATTTGATTAACATGTCAGCCGCTCAAACTCAAC 4682
 QY 211 nLysGluAlaIleTyraAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGlu-P 231
 DB 4681 CAAGTTTGCTGCCAATTAACAGACAACTTCTCACTAAACGCGGATGCTCTCAGATC 4622
 QY 231 fo-----TyrlleThrThr----- 235
 DB 4621 CACCATCAAAATGATATATTGCTGATCTTTGTCACACGATTATTCACCTAATCC 4562
 QY 236 -----PhePheAsnGluValLeuMetLeuGlyLys-----ThrSerIleSera 250
 DB 4561 CTGATCTACACTCTTTTAAACGCTGCTGTGATGATGATGATGATGATGATGATGATG 4503
 QY 250 spleuSerGluHisValPheAspLeuIleLeuGluLeuTyraAsnIleAspSerHisLeu 270
 DB 4502 --ATTACAAACAAATCTATGATATCATTTTACAGAGCTTAACCCATCAATGCAAGCTGT 4445
 QY 270 euleuSerValIleProGluLeuGluIlePheLysSerAsnAspAspGluIleArgL 290
 DB 4444 TGTGTTAGTGCTACCTGACCTGAGACAAACAGTCTGTCAGAGACGATGCTGAGAGGC 4385
 QY 290 eu----- 290
 DB 4384 TAAGTAGTAAATTATTCAGCATGTACCTCTATTGTCATATTAAATATTTTCTGTTA 4325
 QY 291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAla 309
 DB 4324 CTACAGAGGCAACCACTCTCTGTGCTGTGATGTTCTCCGAAAGAGATGCTCAACTGACGA 4265
 QY 309 erGlnAsnLysProLeuIleTyraGluCysTyriLeuGluArgPheAsnAspIleHisValPro 329
 DB 4264 AAAAGTACCCCAACCTGTGATAAATGTTCTTGGCGGCTTCTGACATCAGCAGCAAC 4205
 QY 329 leArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAla 349
 DB 4204 TTGCGATCAAGTGTGTTCACTTCATGCAATGCACTTCTGCTCAATCACCACCAAGCTTCAGC 4145
 QY 349 ysAspLeuThrGluTyLeuLysValArgSerHisAspProGluGluAlaIleArgHis 369
 DB 4144 ACGATATCACCGAAATTTGCGGCTTGAATATACGATCTTTCAGAGATGTCGCGCACG 4085
 QY 369 spValIleValSerIleValThrAlaLysLysAspIleLeuLeuValAspAsp--- 387
 DB 4084 AGGTGTAAATGCTATTGTGGAACATGCAAGGCGCATCTCACCTCTGTTGGAAGCGC 4025
 QY 388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTyrArgValArgLysG 407

```

Db      4024  CCGATCTACTAGAAATTTGGCGGAGCGAGCGCTAGATAAGAAATTCACAGG 3965
Oy      407   luAlaMetMetGlyLeuAlaGlnIleTyrlsYsTyrlAlaLeuInserAlaAlaGlyL 427
Db      3964  ATGCATTAATGATGTGGCTTACATCTACAAAGCGGCAATTTGGCAACCAAGATCTAA 3905
Oy      427   yAspAlaAlaLys---GlnIleAlaTrpIleYAspLysLeuLeuHisIleTyrlYrg 446
Db      3904  GCACGCGCTTAAGGTACAGGCTGACTGATTAAACAAGATACATGATGATCTACA 3845
Oy      446  lnaSerSerIleAspAlaTrpLeuLeuValGluArgIlePheAlaGlnTyrlMetValProH 466
Db      3844  AAGTGGCTTGAGAGCGCGCTGTGGAAAGCGCTACTATACCTGCGCTGTTCCT 3785
Oy      466  lsaSerLeuGlnThrTrpGluArgMetLysCysLeuTyrlTyrlLeuTyrlAlaThrLeuAspL 486
Db      3784  ATAACTAGTCCCGAAGAGCGCATGAAGCTATATCATTTGCTAGCGCATTCGATG 3725
Oy      486  euAsnAlaValLysAlaLeuAsnGlnMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
Db      3724  CCATGCCACCAAGGCTTTCGTCGAGCTGCAGAGAACCAATGAAGACCGCAATACGG 3665
Oy      506  aLysAspLeuLeuAspLeuIleLysGln-----ProLysThrAspAlaSerV 522
Db      3664  TCAGTGTATGTGATCAAGCTACACACCTAAGAGATTACGCGCGGCTAGTACTAGC 3605
Oy      522  aLysAla----- 524
Db      3604  TCACGCGCCAGAGCCCAACATTCGCCAAGTGGTTCTAATTAAGTTTCTTTGATAGT 3545
Oy      525  --IlePheSerLysValMetValIleThr--ArgAsnLeuProAspProGlyLysAlaGln 543
Db      3544  TTGTAACACAGAGATGTGTTTTTTTCTTCAGACGCTCCAGATCCCTTAAGACGGCA 3485
Oy      544  AspPheMetLysLysPheThrGlnValLeuGlnAspAspGlyLysIleArgLysGlnLeu 563
Db      3484  GAGTACCTTACCCAGTTTACACAACTCGGAAAGATGCTCACGCTCTACAGATGATC 3425
Oy      564  GluValLeuValSerProThrCysSerCysLysGlnAlaGlnGlyCysValArgLysIle 583
Db      3424  AACATTCCTCTTAAACGCGAGCTAGAGCTGTGGGAGTGTGCTGACACAGATGGAGTTC 3365
Oy      584  ThrLysLysLeuGlnLysAsnProLysGlnProThrAsnProPheLeuGlnMetIleLysPhe 603
Db      3364  CTGAAAGAGCTTGGCGCCACAGTCCA---TCGAATCTGTATTAACAACAGATTAAAGTG 3308
Oy      604  LeuLeuGlnArgIleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLys 623
Db      3307  CTGATTGAGCGCGTGCATGATGTGTGCAAGAGAGTCTATTTGGCTACTTATAGG 3248
Oy      624  -----Gln 624
Db      3247  TAAATGCTTTTAAGTCCCTTTTGGCAGACTTAATCCGTAATCTTTTGTCTCAGCTTA 3188
Oy      625  ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspLys 644
Db      3187  ATTAAACAATGCAATCGAAGAAAGGTCATGTGTGAAGAGATCGGAATTTGCGTCAAGAA 3128
Oy      645  AlaIleArgAlaGlyLeuGlnLeuLeu----- 653
Db      3127  GCGGGCGAGCGCGGAGCTCAATGTTGTGGCCCTGAGATAGAGATTAATTAAGTATAGTA 3068
Oy      654  -----LysValLeuSerPheThrHisProLysSer 663
Db      3067  CAGAAATCCATTTGATGCTTTATCTTATCTTGCAGATG--CTTTCATAGCTTTTGGCTCAT 3009
Oy      664  PheHisSerAlaGluThrPheGlnSerLeuLeuAlaCysLeuLysMetAspAspGlyLys 683
Db      3008  TTCTTCACCGATCTGCTGCGTCATTAATTTCCCTGCTCAGTTACAGCAGGATTAAT 2949
Oy      684  ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly-----SerLysIleGlu 700

```

```

Db      2948  GTTGACACACTGTGCTCTAAAGACACTAATCATCTGTGGGCGCTATCAGCCACTATGAC 2889
Oy      701  GluAspPheProHisIleArgSerAlaLeuLeuProValIleuHisIleLysSerLysLys 720
Db      2888  GACCTTACACCGGCTATTCTGCAGAGGTGGCTCGGTGTGCAGAGACTTGCACATTAT 2829
Oy      721  GlyProProArgGlnAlaLysTyrlAlaIleHisCysIleHis----- 734
Db      2828  GGAACCCCAAGCAAGCAAGACACCGGCTGCATATTATTTGCTAAACAGCAGTCGTCG 2769
Oy      735  -----AlaIlePheSerSerLysGlnThrGln----- 743
Db      2768  GCTTCACCTAGTATGAGCAACAAGTGCAGATGCGATGCGTCCACAACTAACCAAGATG 2709
Oy      744  -----PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db      2708  CATCTTATTTTCAACAGATCATGCAAGAGCTTCGCTCAAACTGACACA---AACGT 2652
Oy      761  GluHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuAlaProAspGln 780
Db      2651  GAGCATCAGCCGACAAAGATTTGTGACTTGGGTGCATTTGCTACAAATATGCCACAGGCC 2592
Oy      781  PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
Db      2591  TTCTTAACGCCCATTAAGAAATATGATGTCGAGCAATTTGTCAAGACCTGCTATCCAA 2532
Oy      801  AspArgLeuProGlyLysLysThrThrLysLeu-----TrpValProAsp 815
Db      2531  GAA---GTTCCGCGCGAGAGACTACGAACTGCGCAGAGACAGTGAATGTGTGCCCAA 2475
Oy      816  GluGluValSerProGlyThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
Db      2474  GAGAACTACCGCGCGAGACACTATGTAAGCTGAGTGGCTCAAGGCTATGCGCCAGTGG 2415
Oy      836  LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 855
Db      2414  CTTTGGGATGTCGTTACCCGAT---GAGCAGCGTGGCCAGAAACATTCGGAATGTGGCG 2358
Oy      856  ThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMet 875
Db      2357  GCGTTTGTTAACCAACAGAGGATTTGCTCGGCCAGAACCGTCTTGGCGGTGCCGAGAA 2298
Oy      876  SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrl 895
Db      2297  TCTTGGCGCGCGCTGGGGGAGCGTGGCCCATGCTCAAGTGTGGAGCAAAAGGCGCTA 2238
Oy      896  HisGluIleIleThrLeuGlnGlnTyrl---GlnLeuCysAlaLeuAlaIleAsnAspGlu 914
Db      2237  GGTGATCAGTACAGCGCTGACAGATTTTGCAGAGCTTCCAGCTGATGTTGATGAGAGC 2178
Oy      914  ----- 914
Db      2177  ATTCATCTTGTGCTTAGAACTTAGTCTCAATGTCTATCTCGGTGTTCCCGTAGCGTATC 2118
Oy      915  ---CysTyrlGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
Db      2117  CGGTGCCA--GAAGTTGGGAAATCTTGTCTCGCAAGCTGCACAAAGATTTAAGCAGAGT 2059
Oy      934  Arg-----LeuProLeuGlnIlyMetAlaIleCysAlaLeuCysAla----- 947
Db      2058  TTGCCACGAACTGTTTCCGCTGACTTCAATGGCTGTGTATGTCTGCGTGGTGTAGAG 1999
Oy      948  -----LysAspProValLysGluArgAlaAlaHisAlaArg--GlnCysLeuValLys 964
Db      1998  ACTGAGAGAGAGTAGTATGTTCAAGAAACCATTCGATCCGTTGCGAGAAAGTATGATAAA 1939
Oy      964  s-----AsnIleAsnValArg 969
Db      1938  TATTTTCAACAGATTTGCAAGACCTTGTGCGTCACTATGACAGAAACGGATTTAAACAAG 1879
Oy      969  GArgGluTyrlLeuLys----- 974
Db      1878  GCGGGAATATCTCAAGACTGTGCGTATGACATGTAGAGTGTTTACTTTTGGAAATCATATT 1819

```



```

OY 797 LeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuThrValProAspGlu 816
Db 106 -----ACAGGTGAAAAAGAAATGGAAGAACTGTGGCTCCAGATGA 144
OY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeu 836
Db 145 GAGGTTCCCTCGAAGTACTAGCAAGAGTACAGCAATTAACCTCTGGTAAGTGGCTG 204
OY 837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThr 856
Db 205 TTGGGTATGAAAAAACACAGCTGTAAATCTGCCAATTCACACCTTCGGTTATTATCAGCG 264
OY 857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
Db 265 ATGTTGGTTAGTGAAGGGTGAACCTGACAGACGAAAAAGAGATCAGTAATCTATATGTCT 324
OY 877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysThrHis 896
Db 325 CGCTGGCATTAAGTGTGTGTGTCATTAATGAAGCTTGCTCAGGAACCTTGTTACCAT 384
OY 897 GluIleIleThrLeuGluGlnThrGlnLeuCysAlaLeuAlaIleAsnAspGluCysThr 916
Db 385 GAAATTTATTACCCCAAGACAGTTTCAGCTGTGTGCATTTATTAAATGATGAGTGTAC 444
OY 917 GluValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
Db 445 CAAGTAAGCGCATATTTGCTCAGAAAGCTGCATAAGGCACTTGTAAGTACTGCTCCCA 504
OY 937 LeuGluThrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
Db 505 TTGGCATATATGGCATCTTTGCTGTGCCAAGATCTGTGAAGAGAGAGAGA 564
OY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgGluThrLeuLysGlnHis 976
Db 565 CACGCGACGACATGTTTACGAAAAATATCACTATACCCAGGAAATTAATTAAGCAGAT 624
OY 977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluThrValValProThrThrIle 996
Db 625 CCTATGGCTACTGAGAAATTTATTTACTGTTGCTGATATATGTGTTCCATACATGAT 684
OY 997 HisLeuLeuAlaHisAspProAspThrValLysValGlnAspIleGluGlnLeuLysAsp 1016
Db 685 CACGCGTACGCCCATGATCCAGATTTTACAGATCACAAAGATGTTGATCAGCTTCGTGAT 744
OY 1017 ValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsnGluAsnSer 1036
Db 745 ATCAAAAGAGTCCATATGTTGCTCATGCTGAAGTTTAAAGACAAAGAAATGAACAATAGC 804
OY 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
Db 805 CATGCCCTTTATGAAGAGATGCGCAGAGACATCAAGTTAAACGAGATGCCAGTCTCCA 864
OY 1057 AspAspAlaLysMetAsnGluLysLeuThrValLysAspValAlaMetAsnIleIle 1076
Db 865 GATGATCTCAAGACAAATGAATAAACTGTATACAGATGATGATGCTCTCTGTGTATA 924
OY 1077 MetSerLysSerThrThrLysSerLeuGluSerProLysAspProValLeuProAlaArg 1096
Db 925 AATAGTAAAGTCTTGTGT- :CAATTCGACATTTCCAAAAGCAGCCCTCCCAATGAAT 982
OY 1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnThrLysLeuProGluMet 1116
Db 983 TTTTACACA---CCGAAAGAGACTCTGTAAACGATTAAGAGTTATTTTACAGACAGACA 1039
OY 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
Db 1040 AGAGACTTCTGTAAACAGAAAGCAAGCCCTGAGTACAGTACAGTAAATAAG 1099
OY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
Db 1100 CTTTATATCGAAGCGGAAGAACCCCTATGTTTAAACACCTGCGCAGTACGAGTGAAG 1159
OY 1157 -----AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg--- 1173

```

```

Db 1160 AATATTAAATGTAAATTCAGAGCTGAACCTTCAACCGGAATGATCAAGGACAGAGT 1219
OY 1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp-----TyrThrMetSerSer 1191
Db 1220 TCAGAGCGACGCGAAGAACTGAGTGTGTAATAAGAGAACCTGTGAGATTTATTCA 1279
OY 1192 ProLeuProGlyLysLysSerAspLysThrAspAspSerAspLeuValArgSerGluLeu 1211
Db 1280 GTCACACCTGTAAACAAATATTGACCCAGTAAAGAAATAG-----GTAAATAATGCAATTG 1333
OY 1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGlu-GluLysLeuGlyLys 1231
Db 1334 CAAGGAGGAAATATGAAAGCCCAAGACAGAGCGCTCCAGCTTCGCAAAAACCTTGAT 1393
OY 1231 TasAspLeu-----ThrLysLeuValGlnGln-----Lys 1242
Db 1394 CACAATATGCTGAAACAGAAATGAAGCTCACTTCAGAACACACACTCTGCTTGAA 1453
OY 1242 sProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu 1262
Db 1454 AACTAAAGAGACTATTAATCTCTTTTCACATGACACACAGTCTCTGATGAAATGTACA 1513
OY 1262 nGlnThrProGluGluLysArgLys 1271
Db 1514 GCAGAAACTCTGACAGAGAGGCTTAAA 1541

RESULT 12
ABS56684
ID ABS56684 standard; cDNA; 1104 BP.
XX
AC ABS56684;
XX
DT 23-JAN-2003 (first entry)
XX
DE Euchromosome fragile intelligence delay syndrome protein 20.02 cDNA.
XX
KW Euchromosome fragile intelligence delay syndrome protein 20.02; human;
XX euChromosome fragile intelligence delay syndrome; dementia; gene; ss.
XX
OS Homo sapiens.
XX
OS
XX
PH Key location/Qualifiers
XX CDS 184..732
XX FT /*tag= a
XX FT /product= "euchromosome fragile intelligence delay
XX syndrome-associated protein, 20.02"
XX
XX
XX CN1351041-A.
XX
XX 29-MAY-2002.
XX
XX 26-OCT-2000; 2000CN-0125797.
XX PF
XX 26-OCT-2000; 2000CN-0125797.
XX PR
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
XX WPI: 2002-637117/69.
XX P-PSDB; ABB84560.
XX
XX Human euchromosome fragile intelligence delay syndrome associated
XX protein 20.02 polypeptide, used to treat e.g. dementia -
XX
XX Claim 6; Page 29-30 (disclosure); 33p; Chinese.
XX
XX This invention describes the novel human euchromosome fragile
XX intelligence delay syndrome associated protein, 20.02. The polypeptide is
XX used in treating diseases such as euchromosome fragile intelligence delay
XX syndrome, and dementia. This sequence encodes the human euchromosome
XX fragile intelligence delay syndrome-associated protein, 20.02 described

```

CC In the disclosure of the invention.

XX
SQ Sequence 1104 BP; 362 A; 188 C; 228 G; 326 T; 0 other;

Alignment Scores:

Pred. No.:	3,13e-58	Length:	1104
Score:	967.00	Matches:	183
Percent Similarity:	90.56%	Conservative:	28
Best Local Similarity:	78.54%	Mismatches:	22
Query Match:	13.44%	Indels:	1
DB:	24	Gaps:	0

US-09-512-581B-2 (1-1391) x ABB56684 (1-1104)

```
QY      348  AAlaLysAspLeuThrgLurLLeuLysValArgSerHisAspProGluGluAlaLeuArg 367
Db      2    GCGAAGATCTCAGAGATATTAAAGGTAGATCAGATATCCAGAGAGAGCATTCGT 61
QY      368  HIsAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp 387
Db      62  CAGGATGTCATTGTTACTATATATACAGCTGCCAGGAGGAGCCTGCC-TTAGTAATATGAT 120
QY      388  HIsLeuAsnBpheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGlu 407
Db      121  CACCTGCTTGCTTTGTAAGGGAAGACAGCTGATTAACGGTGGCAGTAAGAAAGAA 180
QY      408  AlameLwecLysLeuAlaGlnIleLysLysLysTyraLeuGlnSerAlaAlaGlyLys 427
Db      181  GCTATGATGGCTGCTGGCTCAGCTTTATAGAAATACGTTTCATGCTGTAAGCAGGAAG 240
QY      428  AspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleLysLysGlnAsn 447
Db      241  GAAGCTGCAGAGAAAGACAGCTGATTAAGACAACTTCGATATATTATTCAGAAC 300
QY      448  SerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnIleLysMetValProHisAsn 467
Db      301  AGCATGACACAAATTTGTTGTAAGAAATCTTGCTCAGTATCTTGCCCCACAC 360
QY      468  LeuGluThrThrgLurGluArgMetLysCysLeuLysLysLysLysLysLysLysLysLys 487
Db      361  CTGGAACAGACAGAGAAATGAATGCTTATATTACTTATATGCTGATTTGGATCCAAAT 420
QY      488  AlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLys 507
Db      421  GCGTAAACCTCTCAACGAAATGTGACAACTGTCCGACCACTGTACGC 480
QY      508  AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527
Db      481  GAACATATTGATTTGCACAAAGCAGCTACATCAAGAGCTTACATGTTCTGCATGTTTGA 540
QY      528  LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
Db      541  AAATGATGACATACAGAAATTTGCTGACCCCGGAAAGCACAAAGATTTTGTGAAG 600
QY      548  LysPheThrGlnIleuGluAspArgLysLysIleArgLysGlnLeuGluValLeuVal 567
Db      601  AAATTAACCAAGCTTCGCGCATGTAGAAACTTCGCTCAGATTGGAGTTATATT 660
QY      568  SerProThrcysSerCysLysGlnAlaGluGlyCysVal 580
Db      661  AGCCCAACTGTTCTTGACACACAGCATATTGTGTG 699
```

RESULT 13
AAK85948
ID AAK85948 standard; DNA: 7943 BP.
XX AAK85948;
AC
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40760.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW

```
KW      cytosolic; gene therapy; vaccine; metastasis; ds.  
XX  
XX      Homo sapiens.  
XX  
PN      WO200157182-A2.  
XX  
XX      09-AUG-2001.  
XX  
XX      17-JAN-2001; 2001WO-US01354.  
PR      31-JAN-2000; 2000US-0179065.  
PR      04-FEB-2000; 2000US-0180628.  
PR      24-FEB-2000; 2000US-0184664.  
PR      02-MAR-2000; 2000US-0186350.  
PR      16-MAR-2000; 2000US-0189874.  
PR      17-MAR-2000; 2000US-0190076.  
PR      18-APR-2000; 2000US-0198123.  
PR      19-MAY-2000; 2000US-0205515.  
PR      07-JUN-2000; 2000US-0209467.  
PR      28-JUN-2000; 2000US-0214886.  
PR      30-JUN-2000; 2000US-0215135.  
PR      07-JUL-2000; 2000US-0216647.  
PR      11-JUL-2000; 2000US-0217487.  
PR      11-JUL-2000; 2000US-0217496.  
PR      14-JUL-2000; 2000US-0218290.  
PR      26-JUL-2000; 2000US-0220963.  
PR      26-JUL-2000; 2000US-0220964.  
PR      14-AUG-2000; 2000US-0224518.  
PR      14-AUG-2000; 2000US-0224519.  
PR      14-AUG-2000; 2000US-0225213.  
PR      14-AUG-2000; 2000US-0225214.  
PR      14-AUG-2000; 2000US-0225266.  
PR      14-AUG-2000; 2000US-0225267.  
PR      14-AUG-2000; 2000US-0225268.  
PR      14-AUG-2000; 2000US-0225270.  
PR      14-AUG-2000; 2000US-0225447.  
PR      14-AUG-2000; 2000US-0225757.  
PR      14-AUG-2000; 2000US-0225758.  
PR      14-AUG-2000; 2000US-0225759.  
PR      18-AUG-2000; 2000US-0226279.  
PR      22-AUG-2000; 2000US-0226681.  
PR      22-AUG-2000; 2000US-0226688.  
PR      22-AUG-2000; 2000US-0227182.  
PR      23-AUG-2000; 2000US-0227009.  
PR      30-AUG-2000; 2000US-0228924.  
PR      01-SEP-2000; 2000US-0229287.  
PR      01-SEP-2000; 2000US-0229343.  
PR      01-SEP-2000; 2000US-0229344.  
PR      01-SEP-2000; 2000US-0229345.  
PR      05-SEP-2000; 2000US-0229509.  
PR      05-SEP-2000; 2000US-0229513.  
PR      06-SEP-2000; 2000US-0230437.  
PR      06-SEP-2000; 2000US-0230438.  
PR      08-SEP-2000; 2000US-0231242.  
PR      08-SEP-2000; 2000US-0231243.  
PR      08-SEP-2000; 2000US-0231244.  
PR      08-SEP-2000; 2000US-0231413.  
PR      08-SEP-2000; 2000US-0231414.  
PR      08-SEP-2000; 2000US-0232080.  
PR      08-SEP-2000; 2000US-0232081.  
PR      12-SEP-2000; 2000US-0231968.  
PR      14-SEP-2000; 2000US-0232397.  
PR      14-SEP-2000; 2000US-0232398.  
PR      14-SEP-2000; 2000US-0232399.  
PR      14-SEP-2000; 2000US-0232399.  
PR      14-SEP-2000; 2000US-0232400.  
PR      14-SEP-2000; 2000US-0232401.  
PR      14-SEP-2000; 2000US-0233063.  
PR      14-SEP-2000; 2000US-0233064.  
PR      14-SEP-2000; 2000US-0233065.  
PR      21-SEP-2000; 2000US-0234223.  
PR      21-SEP-2000; 2000US-0234274.  
PR      25-SEP-2000; 2000US-0234997.
```


PR 25-SEP-2000; 2000US-0234988.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure: SEQ ID NO 40760; 3071bp + Sequence Listing: English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK4703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7943 BP: 2567 A; 1175 C; 1468 G; 2733 T; 0 other;

Alignment Scores:
Pred. No.: 2,6e-53 Length: 7943
Score: 912.50 Matches: 180
Percent Similarity: 85.31% Conservative: 0
Best Local Similarity: 85.31% Mismatches: 0
Query Match: 12.69% Indels: 31
DB: Gaps: 1

US-09-512-581b-2 (1-1391) x AAK85948 (1-7943)

QY 1212 GUAUAGGCTAGAGCGAGGAAAAAAGCCGCTCAGAAACGAGAGCAAAATTAAGTATG 61
DB 2 GAGAAAGCTAGAGCGAGGAAAAAAGCCGCTCAGAAACGAGAGCAAAATTAAGTATG 61

QY 1232 ASPASPLEUTHRLYSLEUVALGInGluInLysProLysGlySerGlnArgSerArgLys 1251
DB 62 GATGACTTGACTAAGTTGTCAGAGAACAGAACTAAAGGACGTCACGAAAGTGGGAAA 121

QY 1252 ARGGLYHISThrAlaSerGluSerAspGluGlnGlnIntPrProGluGluLysArgLeuLys 1271
DB 122 AGAGCCCATACGGCTTCAGAAATCTGATGAACAGCAGTGGCTGAGGAAAAAGAGCTCAA 181

QY 1272 GUAUASPLEUAGUASInGluASpGluInLysnsrPrProLysGlyLysArgLys 1291
DB 182 GAAATATATTAAGAAATGAAGATGAACAGATGATGCGCCAAAGAGGTAAGAGAGGC 241

QY 1292 ARGProPolysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
DB 242 CGACACCAAAAACCTCTTGGTGGAGGTACACCAAAAGAGGCAACATGAAACTCTCT 301

QY 1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluGluGluGluGlu 1331
DB 302 AAAAAAGAGACCAAAAAAATCTGGACCTCCAGACCAAGAGGAGGAGAAAGAAAGAA 361

QY 1332 ARGGInserGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArg 1351
DB 362 AGACAAAGTGAATATACGAAACAGAGTCCAAAGCAACACACCGAGTGTCAAGAGGA 421
QY 1352 AlaGlnGln----- 1354

D 422 GCACAGCAGAGTACGATGTACTTAACTGCTGTTTCTACTATATTTAA 481
Q 1355 -----ArgalagIuserProglu 1360
D 482 ATCAATATTGATGCTATCCACATTTGGTCTTCCCAAGACAGAGATCTCCGAA 541
Q 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnIysGlyArgGlyArgProSerLys 1380
D 542 TCTAGTGCATTTGATTCACACACAGCTCCACACAGAAAGAGCGAGAGACCATCAAAA 601
Q 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
D 602 ACGCATCACCATTCACACCAAAAAAATGTG 634
RESULT 14
AAS44727
ID AAS44727 standard; DNA: 714 BP.
AC AAS44727;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #152.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytoskeletal; antirheumatic; antiarthritic; vulnerrary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiParkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiallergic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN W0200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WC-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSED INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI: 2001-589862/66.
DR P-PSDB; AAU27827.
XX
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
XX
PS Claim 1; SEQ ID NO 152; 153bp; English.
XX
XX
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system

CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/publshd_pcl_sequences](http://wipo.int/pub/publshd_pcl_sequences).
XX
SQ Sequence 714 BP; 293 A; 131 C; 178 G; 112 T; 0 other;
XX
Alignment Scores:
Pred. No.: 2.68e-51 Length: 714
Score: 864.00 Matches: 166
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.01% Indels: 0
DB: 22 Gaps: 0
US-09-512-581B-2 (1-1391) x AAS44727 (1-714)
Q 1226 GlnGluLysLeuGlyMetAspPleuThrLysLeuValGlnGlnGlnLysProLysGly 1245
D 3 GAGGAGAAATTAGTATGATGACTTGACTTAAAGTGTACAGGAAACGAACTTAAAGGC 62
Q 1246 SerGlnArgSerArgLysArgGlyHisThiAlaSerGluSerAspGlnGlnIntPro 1265
D 63 AGTCAGCGAAGTCGGAAGAGAGCGCATACGCTTCAGAAATCTGATGAACGACGTGGCCT 122
Q 1266 GlnGluLysArgLeuLysGluAspIleLeuGlnLysGlnLysGlnAsnSerProPro 1285
D 123 GAGGAAAGAGGCTCAAGAAAGATATATTGAAATGCAATGAAAGAAATAGTCCGCCA 182
Q 1286 LysLysGlyLysArgGlyArgProProLysProLeuGlyGlyThrProLysGluGlu 1305
D 183 AAAAAGGCTAAAGAGGCGCACCAACCAAACTCTTGCTGAGGTACACCAAAAGAAAG 242
Q 1306 ProThrMetLysThrSerLysLysGlySerLysLysSerGlyProProAlaProGlu 1325
D 243 CCAACATGAAACTTCTTAATAAAGAAAGCAAAAAAAATCTGACCTCCAGCACAGAG 302
Q 1326 GlnGluGlnGluGlnGluArgGlnSerGlyAsnThrGlnGlnLysSerLysGln 1345
D 303 GAGGAGGAAAGAAAGAAAGCAAAATGCAATACGGAACGAAAGTCCAAAAACCAACAG 362
Q 1346 HisArgValSerArgAlaGlnGlnGlnArgAlaGlnSerProGlnSerSerAlaIleGlu 1365
D 363 CACCGAGTGTCCAAAGAGAGACACAGACAGAGCGAATCTCTTAATCTAGTGCATTTGAA 422
Q 1366 SerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSer 1385
D 423 TCCACAGATGCCACACACAGAAAGCGAGAAAGACATCAAAACGCCATACACATCA 482
Q 1386 GlnProLysLysAsnVal 1391
D 483 CAACCAAAAAAATGTG 500
RESULT 15
AAF18338
ID AAF18338 standard; DNA: 1217 BP.
XX
AC AAF18338;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 357.
XX

Human; lung cancer associated protein; neuroprotective; cytostatic; cardiocactive; immunomodulatory; muscular active; vulnerary; gastrointestinal; nephrotropic; antiinfective; gynecological; antibacterial; diagnosis; neural disorder; immune disorder; reproductive; proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.

PN MO20005180-A2.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05918.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI (ROSE/) ROSEN C A.

PI Ruben SM;

DR WPI: 2000-587514/55.

DR P-PSDB: AAB58462.

PT Lung cancer associated gene sequences, referred to as lung cancer antigens, useful for treatment, prevention, and diagnosis of disorders such as lung cancer -

PS Claim 1; Page 815; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer associated proteins represented in AAB58106 - AAB58548. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active general; vulnerary; gastrointestinal general; nephrotropic; antiinfective; gynecological; or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The proteins may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders. The proteins may also be used in the treatment of wounds and infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are used in the course of the invention for the identification and characterisation of the polynucleotide and protein sequences.

CC CC

XX Sequence 1217 BP; 445 A; 232 C; 270 G; 263 T; 7 other;

SO

Alignment Scores:

Pred. No.: 6,38e-44 Length: 1217

Score: 762.50 Matches: 185

Percent Similarity: 56.88% Conservative: 63

Best Local Similarity: 42.43% Mismatches: 121

Query Match: 10.60% Indels: 68

DB: 21 Gaps: 15

US-09-512-581B-2 (1-1391) x AAF18338 (1-1217)

QY 965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984

DB 2 AATATACAGTATACGAGGATACATTAACAGAAATCTATGGCTACGACAAATTTATTA 61

QY 985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004

DB 62 TCACGTGTGCTGATATAGTATGTCATACATGATTCACCTGCTAGCCCATGATCCAGAT 121

QY 1005 TyrValLysValGlnAspIleGluLeuLysAspValLysGluCysLeuTyrPheVal 1024

DB 122 TTACAGATCATCAGATGTTGATCAGCTGCTGATATCAAGAGTGGCTATGTTGTCATG 181

QY 1025 LeuGluIleuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044

DB 182 CTTGAAGTTTAAATGACAAAGATGAAATGCAATGCGCTTATGACAAAGATGGCA 241

QY 1045 GluAsnIleLysGlnThrLysAspAlaGlnGluProAspAspAlaLysMetAsnGluLys 1064

DB 242 GAGAAACATCAAGTTAAACGAGATGCCAGCTCCAGATGATCCAAACAAATGCAAAA 301

QY 1065 LeuTyrThrValLysAspValAlaMetLysIleIleMetSerLysSerThrThrTyrSer 1084

DB 302 CTGATATACAGTATGATGATGCTGCTGCTTTAAATAGTAAAGTCTTTGTGCAT 361

QY 1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104

DB 362 GCAGATTACCAAGAGGACCAAGCTCCCATGAAATTTTACACAACTGAAAGAC 421

QY 1105 PheSerAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThrProGluLys 1124

DB 422 TTCTGTAACGATTAAGATATATTTACAGAAAGACATCTTCTTAAACAGAAAG 481

QY 1125 ProLysThrThrAsnValLeuGluValAlaValAsnLysProLysSerSerAlaGluLysGln 1144

DB 482 CCAAGGCTGCTGAGATCTAGTGACAGTAAATTAACCTTTATACAGCAGGGAAGAAA 541

QY 1145 SerGlnThrLysSerSerArgMetGluThrValSer-----AsnAlaSerSerSer 1162

DB 542 CCTATGTTAAGACAGACAGCAGCTGAGACCTGGAACATATTAATGTAATTCAGAGCTG 601

QY 1163 AsnProSerSerProGluArgIleLysGluVal---LeuAspSerSerGluMetAspHis 1181

DB 602 AACCTTCAACCGGAAATCGATCAAGGACAGGATTCAGAGCAGCAAGAACTGAGATT 661

QY 1182 SerGluAsnGluAsp-----TyrThrMetSerSerProLeuProGluLysLysSerAsp 1199

DB 662 AGTGAAATGAGAGAACCCCTGAGAGATTATTCAGTACACACCTGTTAAACAAATTTGAC 721

QY 1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGluArgLysLys 1219

DB 721 ----- 721

QY 1220 ThrProValThrGluGlnGluGluLysLeuGluMetAspAspLeuThrLysLeuValGln 1239

DB 722 ---CCAGTAAAGAAATAAAGAA-----ATTAAATCTGATCAGGCTACAC----- 760

QY 1240 GluGluLysProLysGluSerGlnArgSerArgLysArgGluHisThrAlaSerGluSer 1259

DB 761 ---CAGGCAACATCAGCAGTGCACCGAGAAAGAAAGAAACAGTAACAGCAGCTGTGCA 817

QY 1260 AspGlu---GlnGlnTyrProGluGluLysArgLeuLysGluAspIleLeuGluAsnGlu 1278

DB 818 GAGATATATCAAAACAAACAAAGATGAGAAA-----CTA 850

QY 1279 AspGluGlnAsnSerPro---ProLysLysGluLysArgGluArgProProLysProLeu 1297

DB 851 GATGAATCGGAGACCTCCGCGCCCTCCAAACCGAGAGAGAGAGCTGCACCAACATCTGAA 910

QY 1298 Gly---GlyGlyThrProLysGluGluProThrMetLysThrSerLysGluSerLys 1316

DB 911 TCTCAGGCAATGCTACCAAAATGATGATCTTAATTAACCTATTAAACAGGGAAGAG 970

QY 1317 LysLysSer---GlyProProAlaProGluGluGlu----- 1338

DB 971 AGACGTGAGTGGGTACAGAGAGAGCCCTGGGGTTCGAGAGGTAATGCCAAAGACACC 1030

QY 1339 ---GluGluGluAlaGlnSerGluLysThrGlnLys-----SerLys 1342

DB 1031 AAACGTCAAGATTT-AGCCAAAGAGGAGCAGCAGCAAGAAAGCAAAATGACTTACAAAG 1089

QY 1343 SerLysGlnHis-----ArgValSerArgArgLagGlnArgLagLysSerProGlu 1360

DB 1090 GTRAAATGCAATTTTGCAAGAGGAGAAATGAAGGCCAAGCAAGAGCTGCCAGTTTN 1149

QY 1361 SerSerAlaIle-----GluSerThrGlnSerThr 1370

Db 1150 TGC AAAA NCTTGGATTACAATGKCTGACAGAAATGACTTATTCAACC 1197

Search completed: September 24, 2003, 20:46:25
Job time : 949 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 19:50:39 : Search time 175 seconds
(without alignments)
2051.150 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193
Sequence: 1 MAHSKTRTNDGKITPPGVK.....QKGGRPSKTPSPQPKKNV 1391

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7155	99.5	1447	4 O9NT15	O9nt15 homo sapien
2	7155	99.5	1483	4 O9Y215	O9y215 homo sapien
3	7101	98.7	1391	4 O9Y451	O9y451 homo sapien
4	4668.5	64.9	1297	4 O8NG14	O8ng14 homo sapien
5	3787	52.6	824	11 O8BLH6	O8blh6 mus musculu
6	3456	48.0	701	11 O8BJ18	O8bj18 mus musculu
7	2698.5	37.5	851	4 O9Y4D4	O9y4d4 homo sapien
8	2517	35.0	529	4 O8IXT6	O8ixt6 homo sapien
9	2487	34.6	479	4 O96KV4	O96kv4 homo sapien
10	2476.5	34.4	536	4 O96KV3	O96kv3 homo sapien
11	2474	34.4	600	4 O8N7J4	O8n7j4 homo sapien
12	2317	32.2	584	11 O8VDS0	O8vds0 mus musculu
13	1982.5	27.6	1218	5 O9V6A6	O9v6a6 drosophila
14	833.5	11.6	1205	3 O9HFF5	O9hff5 sciliosacch
15	808	11.2	1638	10 O9FTL0	O9ftl0 arabidopsis
16	807	11.2	1579	5 O9XTP6	O9xtf6 caenorhabdi

17	791	11.0	198	11 O9D337	O9d337 mus musculu
18	747.5	10.4	1596	3 O9UYV6	O9uyv6 sordaria ma
19	656	9.1	125	4 O9H5N8	O9h5n8 homo sapien
20	640.5	8.9	333	4 O96DB6	O96db6 homo sapien
21	632.5	8.8	1506	3 O94076	O94076 emericella
22	602	8.4	192	11 O8BX77	O8bx77 mus musculu
23	572.5	8.0	1378	10 O8S1M0	O8s1m0 oryza sativ
24	571.5	7.9	1303	10 O9CAP7	O9cap7 arabidopsis
25	487	6.8	415	5 O9SU25	O9su25 drosophila
26	423.5	5.9	1033	5 O8MWT4	O8mwt4 dictyostell
27	309.5	4.3	390	3 O94237	O94237 schizosacch
28	303	4.2	780	10 O9SA10	O9sa10 arabidopsis
29	264	3.7	990	10 O9S9P0	O9s9p0 arabidopsis
30	256	3.6	873	10 O8GUP3	O8gup3 arabidopsis
31	248.5	3.5	2158	4 O9Y673	O9y673 homo sapien
32	242.5	3.4	2265	4 O9Y674	O9y674 homo sapien
33	235	3.3	1135	6 O8H2M3	O8h2m3 canis famli
34	234.5	3.3	852	10 O9SZ55	O9sz55 arabidopsis
35	234.5	3.3	3130	5 O8IDX6	O8idx6 plasmodium
36	232.5	3.2	3130	5 O9BK46	O9bk46 plasmodium
37	231.5	3.2	2253	13 P70012	P70012 xenopus lae
38	230	3.2	2501	5 O9NCW7	O9ncw7 drosophila
39	229	3.2	1723	2 O9JMX8	O9jmx8 helicobacte
40	228.5	3.2	18074	5 O917U4	O917u4 drosophila
41	225	3.1	1437	5 O18033	O18033 caenorhabdi
42	223.5	3.1	4717	3 O94248	O94248 schizosacch
43	223	3.1	2139	5 O07569	O07569 entamoeba h
44	223	3.1	3254	5 O9BK45	O9bk45 plasmodium
45	222.5	3.1	1819	16 O9ZLV0	O9zlv0 helicobacte

ALIGNMENTS

RESULT 1	ID	SEQUENCE FROM N.A.	PRT: 1447 AA.
O9NT15	AC	O9NT15	PRELIMINARY:
O9NT15	AC	O9NT15	PRT: 1447 AA.
DT	01-OCT-2000	(TREMBLrel. 15. Created)	
DT	01-OCT-2000	(TREMBLrel. 15. Last sequence update)	
DT	01-OCT-2002	(TREMBLrel. 22. Last annotation update)	
DE	Hypothetical protein.		
OC	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NC	[1]		
RP	SEQUENCE FROM N.A.		
RA	Rhodes S., Huckle E.;		
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL, AL137201; CAB69911.1; -		
DR	InterPro: IPR000637; AT_hook.		
DR	SMART; SM00384; AT_hook; 1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 1447 AA; 16466 MW; 145C30308EA3EFD5 CRC64;		
Query Match	99.5%; Score 7155; DB 4; Length 1447;		
Best Local Similarity	99.7%; Pred. No. 0;		
Matches 1387; Conservative	0; Mismatches 4; Indels 0; Gaps 0;		
Oy	1 MAHSKTRTNDGKITPPGVKESDKSEKVRRLKMYVTFMDMDSDSEKELYNTLA 60		
Db	1 MAHSKTRTNDGKITPPGVKESDKSEKVRRLKMYVTFMDMDSDSEKELYNTLA 60		
Oy	61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSPDKLKIDFMETTRQLGLEDT 120		
Db	61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSPDKLKIDFMETTRQLGLEDT 120		
Oy	121 KSPQNFYFYLENIAMWKSYNICELEDSNEITOLYRTLFYINNGHOKVMHWVL 180		
Db	121 KSPQNFYFYLENIAMWKSYNICELEDSNEITOLYRTLFYINNGHOKVMHWVL 180		
Oy	181 MSTTCGDTVSOELDPLTVLVNLPVPAHKNLNKQAYDLAKALKRTAQAIEPYITTFPNV 240		

```

|||||
181 MSIIIEGDTVSOELDTVLVNLVPAHKNLNKOAYDLAKALLKRTQAITEPYITTFNFGV 240
QY 241 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPOLERKLSNDNEERLOVYKLLAKMF 300
Db 241 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPOLERKLSNDNEERLOVYKLLAKMF 300
QY 301 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKVRSH 360
Db 301 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKVRSH 360
QY 361 DPEEARIRHDIYVSIYTAAKKDILLVNDHLINFEVRETLDRKMRVRKRAMGLAQIYKKYA 420
Db 361 DPEEARIRHDIYVSIYTAAKKDILLVNDHLINFEVRETLDRKMRVRKRAMGLAQIYKKYA 420
QY 421 LOSAAGKDAKQIAWIKDKLILHYONSIDRLIVERIFAQYVWPHNLETTERMKCLYYL 480
Db 421 LOSAAGKDAKQIAWIKDKLILHYONSIDRLIVERIFAQYVWPHNLETTERMKCLYYL 480
QY 481 YATLDINAVKALNEMKCONLRHQVCKDLIDLIKOPRTDASVKAIFSKVAVITRNLPDPG 540
Db 481 YATLDINAVKALNEMKCONLRHQVCKDLIDLIKOPRTDASVKAIFSKVAVITRNLPDPG 540
QY 541 KAODFMKKFTQVLEDEDEKIRKOLEVLSPTCSCKQAEVREITTKLGNPKOPTNPLEM 600
Db 541 KAODFMKKFTQVLEDEDEKIRKOLEVLSPTCSCKQAEVREITTKLGNPKOPTNPLEM 600
QY 601 IKFLERIAVPHIDTESISALIKOVNKSIDGTADDEDEGPTQOATIRAGIELLKVLSFTH 660
Db 601 IKFLERIAVPHIDTESISALIKOVNKSIDGTADDEDEGPTQOATIRAGIELLKVLSFTH 660
QY 661 PISFHSAEFESLILACIKMDEKVAEALQIFKNFGSKIEDEPHIRISALLPYLHNHKKSK 720
Db 661 PISFHSAEFESLILACIKMDEKVAEALQIFKNFGSKIEDEPHIRISALLPYLHNHKKSK 720
QY 721 GPRROAKYAIHCHIAIFSSKETQFAQIFERLHKSLEDSNIEHLITPLVYTGIALLPDQ 780
Db 721 GPRROAKYAIHCHIAIFSSKETQFAQIFERLHKSLEDSNIEHLITPLVYTGIALLPDQ 780
QY 781 FAAPWKSVAATFIVKULMMDRLPGKTKTAWPDEVESEPTVAVKIOATIMVNRWLLGKM 840
Db 781 FAAPWKSVAATFIVKULMMDRLPGKTKTAWPDEVESEPTVAVKIOATIMVNRWLLGKM 840
QY 841 NNHSGSTSTLRLLTTLHSDDGLEOGKISKPDMSRLRLAAGSALYKLAQEPCHYEIT 900
Db 841 NNHSGSTSTLRLLTTLHSDDGLEOGKISKPDMSRLRLAAGSALYKLAQEPCHYEIT 900
QY 901 LEOYQCALAIINDECTQOVYFAOKLHKLISRLRPLEYATCALCAKDPYKERRAHARQ 960
Db 901 LEOYQCALAIINDECTQOVYFAOKLHKLISRLRPLEYATCALCAKDPYKERRAHARQ 960
QY 961 CLVKNINVRREYIKOHAASEKLLSLPEYVVPYTHLHADDPYVVOIDOLKDYKEC 1020
Db 961 CLVKNINVRREYIKOHAASEKLLSLPEYVVPYTHLHADDPYVVOIDOLKDYKEC 1020
QY 1021 LMFVLETLAKNENNSHAFIRKVENIKQTKAOGPDDAKMEKLYTVCCVANNIIMSKS 1080
Db 1021 LMFVLETLAKNENNSHAFIRKVENIKQTKAOGPDDAKMEKLYTVCCVANNIIMSKS 1080
QY 1081 TTYTSLSPKDPVLPARFTQPDKNFSNTKNYLPPEMKSFTPGPKTTNVLGAVNKLSS 1140
Db 1081 TTYTSLSPKDPVLPARFTQPDKNFSNTKNYLPPEMKSFTPGPKTTNVLGAVNKLSS 1140
QY 1141 AGKOSQTSRMETVNASSSNPSPGRITKGRDSEMDHSEMEDYTMSSPLPGKSKDK 1200
Db 1141 AGKOSQTSRMETVNASSSNPSPGRITKGRDSEMDHSEMEDYTMSSPLPGKSKDK 1200
QY 1201 RDDSIVLRSELEKPRGRKKTPTVEQOEKLGMDLTLKVOBQPKSGORSRKRKHTASESD. 1260
Db 1201 RDDSIVLRSELEKPRGRKKTPTVEQOEKLGMDLTLKVOBQPKSGORSRKRKHTASESD 1260
QY 1261 EQOWPEEKLKEDILENEDEONSPPKKGRRGPRPKPLGGGTPKKEEPLMKTSKSGSKKSG 1320
|||||

```

```

Db 1261 EQOWPEEKLKEDILENEDEONSPPKKGRRGPRPKPLGGGTPEEPTMTSKSGSKKSG 1320
QY 1321 PAPAEEREEBROSGTQEQKSKORHVRRAQOARESPESAISTOSTPOKGRGPRSK 1380
Db 1321 PAPAEEREEBROSGTQEQKSKORHVRRAQOARESPESAISTOSTPOKGRGPRSK 1380
QY 1381 TPSPSQPKKNV 1391
Db 1381 TPSPSQPKKNV 1391

RESULT 2
Q9Y215
ID Q9Y215 PRELIMINARY; PRT: 1483 AA.
AC Q9Y215;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE KIAA0979 protein (Fragment).
GN KIAA0979.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL: AB023196; BAA76823.2; -
DR InterPro: IPR000637; At_hbook.
DR PRINTS: PR00929; ATHOOK.
FT NON_TER
SQ SEQUENCE 1483 AA; 168415 MW; 0B328407FD0BDD74 CRC64;

Query Match 99.5%; Score 7155; DB 4; Length 1483;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1387; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MASHKTRTNDGKITYPGVKEISDKISKEEMVRLKMVYKTFMDMDQDSEEEKLYLNTA 60
Db 37 MASHKTRTNDGKITYPGVKEISDKISKEEMVRLKMVYKTFMDMDQDSEEEKLYLNTA 96
QY 61 LHLASDFLKHGPKDVRLVAVCCGLADIFRIYAEARYTSPDKLIDIFMETTRQLKGLDET 120
Db 97 LHLASDFLKHGPKDVRLVAVCCGLADIFRIYAEARYTSPDKLIDIFMETTRQLKGLDET 156
QY 121 KSPQFNRYEFLLENIMAVSYNCFLEDSNEIFQTLYRTLSVINNGHOKVHMVLDL 180
Db 157 KSPQFNRYEFLLENIMAVSYNCFLEDSNEIFQTLYRTLSVINNGHOKVHMVLDL 216
QY 181 MSIIIEGDTVSOELDTVLVNLVPAHKNLNKOAYDLAKALLKRTQAITEPYITTFNFGV 240
Db 217 MSIIIEGDTVSOELDTVLVNLVPAHKNLNKOAYDLAKALLKRTQAITEPYITTFNFGV 276
QY 241 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPOLERKLSNDNEERLOVYKLLAKMF 300
Db 277 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPOLERKLSNDNEERLOVYKLLAKMF 336
QY 301 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKVRSH 360
Db 337 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKVRSH 396
QY 361 DPEEARIRHDIYVSIYTAAKKDILLVNDHLINFEVRETLDRKMRVRKRAMGLAQIYKKYA 420
Db 397 DPEEARIRHDIYVSIYTAAKKDILLVNDHLINFEVRETLDRKMRVRKRAMGLAQIYKKYA 456
QY 421 LOSAAGKDAKQIAWIKDKLILHYONSIDRLIVERIFAQYVWPHNLETTERMKCLYYL 480
|||||

```

```

Db 457 LOSAAGKDAKQJAWTKDKLHITYQNSIDDLVLRIFRQVWVPIINLETTEKMKLYYL 516
481 YATLDLNAVKALENMKCONLLRHQVKDLIDLKQKPTDASVKAIFSKVWVITRNLPDGG 540
517 YATLDLNAVKALENMKCONLLRHQVKDLIDLKQKPTDASVKAIFSKVWVITRNLPDGG 576
541 KADPFMKFTQVLEDEDEKIRKOLEVLSPTSCSKQABGVREITRTKLGKPNKOTNPLEM 600
577 KADPFMKFTQVLEDEDEKIRKOLEVLSPTSCSKQABGVREITRTKLGKPNKOTNPLEM 636
601 IKFLERLAPVHIDTESISALIKQVKSIDGTADDEGVPTDQARAGLELLKVSFTN 660
637 IKFLERLAPVHIDTESISALIKQVKSIDGTADDEGVPTDQARAGLELLKVSFTN 696
661 PISFHSAEFPESLACLKMDKDEKVAEALQIFKNKTSKIEEDFPHIRSAALLPVLLHKKSK 720
697 PISFHSAEFPESLACLKMDKDEKVAEALQIFKNKTSKIEEDFPHIRSAALLPVLLHKKSK 756
721 GPPROAKVAIHCIAHFSSKETQFAQIFEPRLKSLDPSNLEHLITPLVIGHIALLAPQ 780
757 GPPROAKVAIHCIAHFSSKETQFAQIFEPRLKSLDPSNLEHLITPLVIGHIALLAPQ 816
781 FAAPKMSVATFTVKDLMLNDRLPGKKTTLKLVNPDEEVSPEETVWKIQATIKMVRMLLGK 840
817 FAAPKLSVATFTVKDLMLNDRLPGKKTTLKLVNPDEEVSPEETVWKIQATIKMVRMLLGK 876
841 NNHSSGISTRLTLTLIISDGLTQOGKISKPDMSRLKLAAGSAIVKLAQEPCHYEIT 900
877 NNHSSGISTRLTLTLIISDGLTQOGKISKPDMSRLKLAAGSAIVKLAQEPCHYEIT 936
901 LEOYOLCALAINDECQVQVFAQKLGKLSRLRLPLEMAICALCAKPKVERRNAHQ 960
937 LEOYOLCALAINDECQVQVFAQKLGKLSRLRLPLEMAICALCAKPKVERRNAHQ 996
961 CLVKNINVRREYLKQHAASEKLSILPEYVVPYTIHLAHPDYKVDIEQLKDYKEC 1020
997 CLVKNINVRREYLKQHAASEKLSILPEYVVPYTIHLAHPDYKVDIEQLKDYKEC 1056
1021 LMFVLEILMAKNENSHAFIRKVENIKOTKDAQGDDAKMNEKLTVCDAVANNIIMSKS 1080
1057 LMFVLEILMAKNENSHAFIRKVENIKOTKDAQGDDAKMNEKLTVCDAVANNIIMSKS 1116
1081 TTYTLESPPKDPVLPARFTQPDKNFSNTKNYLPRPEKSPFTPGKPKTTVNLGAVNKLPS 1140
1117 TTYTLESPPKDPVLPARFTQPDKNFSNTKNYLPRPEKSPFTPGKPKTTVNLGAVNKLPS 1176
1141 AKGQOTKSSRMETVSNASSSNPSSPGRIKGRLDSEMDHSEMDYTMSSPLPKKSDK 1200
1177 AKGQOTKSSRMETVSNASSSNPSSPGRIKGRLDSEMDHSEMDYTMSSPLPKKSDK 1236
1201 RDDSLVRSLEKPKGRKKTPTTQOEKLGMDLTUKVOEOKPKGQSRKRGHTASEED 1260
1237 RDDSLVRSLEKPKGRKKTPTTQOEKLGMDLTUKVOEOKPKGQSRKRGHTASEED 1296
1261 EOWMPEERKLKEDILENDEQNSPPKKGKRGPRPLGGTPEKEPTMTSKKSGKKKG 1320
1297 EOWMPEERKLKEDILENDEQNSPPKKGKRGPRPLGGTPEKEPTMTSKKSGKKKG 1356
1321 PPAPEEEREROSQNTQOKSKQHRVSRRAQRAESPESAIESTOSTPOKGRGRPSK 1380
1357 PPAPEEEREROSQNTQOKSKQHRVSRRAQRAESPESAIESTOSTPOKGRGRPSK 1416
1381 TPSPSOPKKNV 1391
1417 TPSPSOPKKNV 1427

```

RESULT 3
 09Y451 PRELIMINARY: PRT: 1391 AA.
 AC 09Y451:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)

```

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Androgen-induced prostate proliferative shutoff associated
DE protein.
GN AS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;
RT "Androgen induced proliferative shutoff in prostate cancer cells.";
RL Proc. Am. Assoc. Cancer Res 37:223-223(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;
RT "Early gene expression during androgen-induced inhibition of
RT proliferation of prostate cancer cells: a new suppressor candidate on
RT chromosome 13, in the BRCA2-Rb1 locus.";
RL J. Steroid Biochem. Mol. Biol. 68:41-45(1999).
DR EMBL: U95825; AAD2134.2; -.
DR InterPro: IPR000637; AT_hook.
DR SMART: SM00384; AT_hook; 1.
SQ SEQUENCE 1391 AA; 158035 MW; F58AEES5AD6D9479 CRC64;

Query Match 98.7%; Score 7101; DB 4; Length 1391;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1377; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 MAHSTRINDGKITTPRGKEISDKISKEMVRLKMNVTMTMDODSEEEKELYLNA 60
DB 1 MAHSTRINDGKITTPRGKEISDKISKEMVRLKMNVTMTMDODSEEEKELYLNA 60
QY 61 LHLASDFLKHGKDVRLVACCLADIFRIVAPRATYSPDKIDPFITQLGLEDT 120
DB 61 LHLASDFLKHGKDVRLVACCLADIFRIVAPRATYSPDKIDPFITQLGLEDT 120
QY 121 KSPQNRXYLLLENIAWVSYNICELEDSNEIFQLYRTLFVYINGHOKVHNMYDL 180
DB 121 KSPQNRXYLLLENIAWVSYNICELEDSNEIFQLYRTLFVYINGHOKVHNMYDL 180
QY 121 KSPQNRXYLLLENIAWVSYNICELEDSNEIFQLYRTLFVYINGHOKVHNMYDL 180
DB 121 KSPQNRXYLLLENIAWVSYNICELEDSNEIFQLYRTLFVYINGHOKVHNMYDL 180
QY 181 MSIIICEGDTVSOELDTVLVNLVPAHKNLKNQAVDLAKALKKRTAQAIEPTTFNVOY 240
DB 181 MSIIICEGDTVSOELDTVLVNLVPAHKNLKNQAVDLAKALKKRTAQAIEPTTFNVOY 240
QY 241 LMLGTSISDLSSEHFDDLLELYNIDSHELLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
DB 241 LMLGTSISDLSSEHFDDLLELYNIDSHELLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
QY 301 GAKDELSAQNPKLMQCYLGRPNDIHVPTRLECVAFASHCLNHPDLAKDLELYKVRSH 360
DB 301 GAKDELSAQNPKLMQCYLGRPNDIHVPTRLECVAFASHCLNHPDLAKDLELYKVRSH 360
QY 361 DPEEAI RHVYIYSITAAKDLVLYNDHLNVREERTLDKRRKRVKKAAMGLAQYKKYA 420
DB 361 DPEEAI RHVYIYSITAAKDLVLYNDHLNVREERTLDKRRKRVKKAAMGLAQYKKYA 420
QY 421 LOSAAGKDAKQJAWTKDKLHITYQNSIDDLVLRIFRQVWVPIINLETTERMKCLYYL 480
DB 421 LOSAAGKDAKQJAWTKDKLHITYQNSIDDLVLRIFRQVWVPIINLETTERMKCLYYL 480
QY 481 YATLDLNAVKALENMKCONLLRHQVKDLIDLKQKPTDASVKAIFSKVWVITRNLPDGG 540
DB 481 YATLDLNAVKALENMKCONLLRHQVKDLIDLKQKPTDASVKAIFSKVWVITRNLPDGG 540
QY 541 KADPFMKFTQVLEDEDEKIRKOLEVLSPTSCSKQABGVREITRTKLGKPNKOTNPLEM 600
DB 541 KADPFMKFTQVLEDEDEKIRKOLEVLSPTSCSKQABGVREITRTKLGKPNKOTNPLEM 600

```


QY 601 IKLELRIAPVHIDTESISALIKOVNKSIDGTADDEGEVPTDQAIRAGLELLKVLSTFH 660
 Db 601 IKFLERLAPVHIDTESISALIKOVNKSIDGTADDEGEVPTDQAIRAGLELLKVLSTFH 660
 QY 661 PISHSAPTESLACLMKDEKVAEALQIFKNTGSKIEEDFPHISALLPVJHHSKK 720
 Db 661 PISHSAPTESLACLMKDEKVAEALQIFKNTGSKIEEDFPHISALLPVJHHSKK 720
 QY 721 GPPQAKYAIHCIAIFSSKETOPAQIFEPHLKSLDPSNLEHLTPVTYIGHIALLPDQ 780
 Db 721 GPPQAKYAIHCIAIFSSKETOPAQIFEPHLKSLDPSNLEHLTPVTYIGHIALLPDQ 780
 QY 781 PAAPWKSVAATFVYKDLMDRLPGKKTTLMPDEEVSPEYMWKIQAIKMMVWLLGMK 840
 Db 781 PAAPWKSVAATFVYKDLMDRLPGKKTTLMPDEEVSPEYMWKIQAIKMMVWLLGMK 840
 QY 841 NNHSGSTSLRLTLTLHSDGDLTEQOKISKPMMSRLAAGSAIYKLAQEPCHYHIT 900
 Db 841 NNHSGSTSLRLTLTLHSDGDLTEQOKISKPMMSRLAAGSAIYKLAQEPCHYHIT 900
 QY 901 LEQYQCALAINDECYOVROVFAOKLHKGSLRLPPEYMAICALCAKADPKERRAHARQ 960
 Db 901 LEQYQCALAINDECYOVROVFAOKLHKGSLRLPPEYMAICALCAKADPKERRAHARQ 960
 QY 961 CLVKNINVRREYKQHAASEKLLSLPEYVVPYTHLLAHDPDYKQVODIEQLKDYEC 1020
 Db 961 CLVKNINVRREYKQHAASEKLLSLPEYVVPYTHLLAHDPDYKQVODIEQLKDYEC 1020
 QY 1021 LMFVLETLMAKNENSHAFIRKQWENTKOTKADGPPDAKNEKLYTVCDVAMNITMSKS 1080
 Db 1021 LMFVLETLMAKNENSHAFIRKQWENTKOTKADGPPDAKNEKLYTVCDVAMNITMSKS 1080
 QY 1081 TTVSLESKDPVLPAREFTOPDKNFSNTKNYLPPEMKSFTTPGKPKTTNLGAVNKPSS 1140
 Db 1081 TTVSLESKDPVLPAREFTOPDKNFSNTKNYLPPEMKSFTTPGKPKTTNLGAVNKPSS 1140
 QY 1141 AGKQSQTKSSMRETVNASSSSNSPGRIRKGLDSSBMDHSEMDYTMSSPLGKRSKSDK 1200
 Db 1141 AGKQSQTKSSMRETVNASSSSNSPGRIRKGLDSSBMDHSEMDYTMSSPLGKRSKSDK 1200
 QY 1201 RDDSGLVRSLEKPRGKKTPTVTEQEEKLGMDLTKLYOEBQPKGSORSRKRGHTASESD 1260
 Db 1201 RDDSGLVRSLEKPRGKKTPTVTEQEEKLGMDLTKLYOEBQPKGSORSRKRGHTASESD 1260
 QY 1261 EQQPEEKRLKEDILENDEQNSPPKKGKRGPRPKPLGGGTPKDEPTMKTSSKSGSKSG 1320
 Db 1261 EQQPEEKRLKEDILENDEQNSPPKKGKRGPRPKPLGGGTPKDEPTMKTSSKSGSKSG 1320
 QY 1321 PPAPEEEDDEERGOGNTQOKSKSQOHVYSRAQAQRAESSAIESTOSTPQKRGPRSK 1380
 Db 1321 PPAPEEEDDEERGOGNTQOKSKSQOHVYSRAQAQRAESSAIESTOSTPQKRGPRSK 1380
 QY 1381 TPSPSQPKKNV 1391
 Db 1381 TPSPSQPKKNV 1391

RESULT 4

Q8NG14 PRELIMINARY; PRT: 1297 AA.
 AC Q8NG14: 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
 DE SCC-112.
 GN SCC-112.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kumar D., Patel S., Whiteside T.L., Kasid U.;

RT "Identification and characterization of SCC-112, a novel cell cycle
 RT regulated gene in human cancer cells."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF294791; AAM82347.1; -
 SQ SEQUENCE 1297 AA; 146610 MW; A42D13E68CA508E CRC64;
 Query Match 64.9%; Score 4668.5; DB 4; Length 1297;
 Best Local Similarity 69.0%; Pred. No. 2.1e-261;
 Matches 908; Conservative 160; Mismatches 199; Indels 49; Gaps 11;
 QY 31 MVRRLKVVVTFMDMODSEBEKLYNLALHLASDFLKHGKDVALLVACCAIDFRI 90
 Db 1 MIKRLKVVVTFMDMODSEBEKQOYLPALHLASEFLRPNKDVALLVACCAIDFRI 60
 QY 91 YAPAPYTPDKLIDIMEFTIROLKGLDEDRSPQNFYFVLLENIAVWKSYNICELEDS 150
 Db 61 YAPAPYTPSHDKLIDILFTIROLKGLDEDRSPQNFYFVLLENIAVWKSYNICELEDC 120
 QY 151 NEIFQYRLFSVYNNHNOKVHMHVMDLMSIICEDGVYQGLDVTVLNVPARKNL 210
 Db 121 NEIFQYRLFSVYNNHNNKKVQMHMLDMSIIMEGDVYQGLDVTVLNVPARKNL 180
 QY 211 NKQAYDLAKALKRTAIEYITTFNOVNLGKTSISDSEHYVFDLIELYNIDSHLL 270
 Db 181 NKQSFDLAKYLLKRTVOTIRACLANFNOVLVGRSSVSDSEHYVFDLIELFAIDPHL 240
 QY 271 LSVLPQLEFKLSNDNEERLOVYKLAHMGAKSELSAONKPLMOCYLGRFNDIHYPIR 330
 Db 241 LSVLPQLEFKLSNDSEGERLAVVRLAKLGSOKDLATONRPLMOCYLGRFNDIHYPIR 300
 QY 331 LECYKFASSHCLMNHDLAKDLTEYLYKVRSHDPEEAIHHDYIVSVITRAKDLILVNDHL 390
 Db 301 LESYKFASSHCLMNHDLAKDLTEYLYKVRSHDPEEAIHHDYIVSVITRAKDLILVNDHL 360
 QY 391 NFVERTLDRMRVRKRAMMGLQIYKKYVLAQSAAGDAKQIAMIKDLIHYONSID 450
 Db 361 GFVAKRTLDRMRVRKRAMMGLQIYKCHGEGAEAEKSWIDKDLIHYONSID 420
 QY 451 DRLVERLIFAQYVPHNLETTERRKCLYUYATLDLNAVALNEMKQNLRLHYOKDL 510
 Db 421 DRLVERLIFAQYVPHNLETTERRKCLYUYATLDLNAVALNEMKQNLRLHYOKDL 480
 QY 511 DLKQKPTDASVKAISFKNVYITRNLPDQKADPMKFFVOVLEDEKIRKOLEVLYSPT 570
 Db 481 DLKQKPTSEANCSAMFCKLMTIAKNLPDPKQADPFVKFFNOVLEDEKIRKOLEVLYSPT 540
 QY 571 CSCQAGCYREITKIKGNKOPNPLMIKFLERITAVVHIDTESISALIKOVNSID 630
 Db 541 CSCQADICVREIARLANRQPTNPLMIKFLERITAVVHIDTESISALIKOVNSID 600
 QY 631 GTADDEGEVPTDQAIRAGLELLKVLSTFHPISFSAETESLACLMDKVAEALQ 690
 Db 601 GTADDEGEVSPDAIRAGLELLKVLSTFHPISFSAETESLACLMDKVAEALQ 660
 QY 691 IFKNTGSKIEEDFPHISALLPVJHHSKSGPPQAKYAIHCIAIFSSKETOPAQIFEP 750
 Db 661 IFKNTGSKIEEDFPHISALLPVJHHSKSGPPQAKYAIHCIAIFSSKETOPAQIFEP 720
 QY 751 LKHSGLDPSNLEHLTPVTYIGHIALLPDQFAAPWKSVAATFVYKDLMDRLPGKKT 810
 Db 721 LKHSGLDPSNLEHLTPVTYIGHIALLPDQFAAPWKSVAATFVYKDLMDRLPGKKT 780
 QY 811 LMPDEEVSPEYMWKIQAIKMMVWLLGMKNHSGSTSLRLTLTLHSDGDLTEQOKI 870
 Db 781 LMPDEEVSPEYMWKIQAIKMMVWLLGMKNHSGSTSLRLTLTLHSDGDLTEQOKI 840
 QY 871 SKPMSRLRLAAGSAIYKLAQEPCHYHITLEQYQCALAINDECYOVROVFAOKLHKG 930
 Db 841 SKPMSRLRLAAGSAIYKLAQEPCHYHITLEQYQCALAINDECYOVROVFAOKLHKG 900
 QY 931 SRLPLPEYMAICALCAKADPKERRAHARQCLVKNINVRREYKQHAASEKLLSLPEY 990
 Db 901 VKLLPLPEYMAICALCAKADPKERRAHARQCLVKNINVRREYKQHAASEKLLSLPEY 960

QY 991 VVPTTHILHADDPRVYKODLEQAKDKVCEKLMFVLEILMAKNENKNSHAFFIKMYENIKOT 1050

Db 961 VVPMIHLHADDPETRSQVDODLRDICEKLMFLEVMYLTKNENKNSHAFFIKMAENIKILT 1020

QY 1051 KDAQPPDDAKNEKLYPCVDVAMNIMSKSTYSLESPKDBVLPAEPFOTPDKAEMTKN 1110

Db 1021 RDAOSPDESKTNEKLYCYDVALCYINSKSALCNADSKDPVLPMPKFTPEQKPCNDKS 1080

QY 1111 YLPEPMKSFETPGKPKTTNYGAVANKPLSSAGKOSKTSRMEYVS--NASSSSNPSPSG 1168

Db 1081 YISEETVLLLTGPKPAGVIGAVANKPLSAGKRPYARSTETEGSNINNNSELPSTGN 1140

QY 1169 RIKGR-LDSEMDHSENE--YTMSSPLPGKSKSDRDSDLVRSELEKPRGRKKTPTYEQ 1225

Db 1141 RSRQSSSEAAETGVSEMEENPVRLIYSTVPKNID-----PYKNK 1179

QY 1226 EEKIGMDLTKLVDEQPKFSQSRKKKGHTASEDE-QQMEERKLKEDILLENDEONSP 1284

Db 1180 E--INSQOAT-----QGNISDSRCKKRTVTMAAGENIIOQKDER-----VDESGPP 1223

QY 1285 -PKKRGGRPKPLG-CGTPEEPTMTSKGSKKSGPARPEEEEBEEROSGUTE 1338

Db 1224 APSKPRKGRRKSESQGNATKNDLNPINKGRG---AAVGGSPGGLGAGNAK 1275

RESULT 5					
ID	Q8BLH6	PRELIMINARY;	PRT:	824 AA.	
AC	Q8BLH6				
Dt	01-MAR-2003 (TREMBlrel. 23, Created)				
Dt3	01-MAR-2003 (TREMBlrel. 23, Last sequence update)				
Dt7	01-MAR-2003 (TREMBlrel. 23, Last annotation update)				
DE	49910.1.2 (Fragment).				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
Ox	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J;				
RX	MEDLINE=22354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	"the RIKEN Genome Exploration Research Group phase I & II Team;				
RT	"Analysis of the mouse transcriptome based on functional annotation of				
RT	60,770 full-length cDNAs."				
RL	Nature 420:563-573(2002).				
DR	EMBL: AK045159; BAC32242.1; -				
FT	NON_TER				
SQ	SEQUENCE	824 AA;	92412 MW;	4FEED73F792C71EC CRC64;	
	Query Match	52.6%;	Score 3787;	DB 11;	Length 824;
	Best Local Similarity	95.3%;	Pred. No. 1,1e-210;		
	Matches 731;	Conservative 13;	Mismatches 23;	Indels 0;	Gaps 0;
OY	625 VNSIDGTADDEDEGCVTDQAIRAGLELVLSFTPIPSFHSATFESLLACLKMDDEVY	684			
Db	1 VNKSIDGTADDEDEGCVPTDQAIRAGLELVLSFTPIPSFHSATFESLLACLKMDDEVY	60			
OY	685 AEALAQTFKNWGSIEEDPFPHISALLPVLIHHSKKKPPPOAKVAIHCHAIINSSKTEOF	744			
Db	61 AEAALQTFKNWGSIEEDPFPHISALLPVLIHHSKKKPPPOAKVAIHCHAIINSSKTEOF	120			
OY	745 AOIFEPLHKLSDPSNEHLITPVTIGIHALLPADOPAAPMKSVWATFIYKDLMDNDLRP	804			
Db	121 AOIFEPLHKLSDPSNEHLITPVTIGIHALLPADOPAAPMLSGMKNHSSGSTLTLLTHSHDGL	180			
OY	805 GKTKTKLMVDEDESPETMWKIOAIKMMVMWLGMKNHSSGSTLTLLTHSHDGL	864			
Db	181 GKTKTKLMVDEDESPETMWKIOAIKMMVMWLGMKNHSSGSTLTLLTHSHDGL	240			
OY	865 TEGGRISKPMRSRLRLAAGSAIVKLAQEPYHEHTITEEQVLQCALAINDECYGVRFQFAQ	924			

Db	241	1E0CKISKPMDSRLRLAAGSAIVKLA0ERCHEIITTELOYQULALAINDECYQR0V9AQ	300
Qy	925	KLHGLSLRLPLEYMAICALCAKDPVKERRAHAROCUVKNIWVREYLK0HAAVSEKLL	984
Db	301	KLHGLSLRLPLEYMAICALCAKDPVERRAHAROCUVKNIWVREYLK0HAAVSEKLL	360
Qy	985	SLREYVVPYTIHLHADPDVYKODIOLMDVRCLEFVLEIIMAKKENNSHAFIRKMY	1044
Db	361	SLREYVVPYTIHLHADPDVYKODIOLMDVRCLEFVLEIIMAKKENNSHAFIRKMY	420
Qy	1045	ENIKOTDAQ0PDDAKMNEKLYTVCDVAMNITMSKSTYVLSLEPKDPLVPARFTQ0PKN	1104
Db	421	ENIKOTDAQ0PDDAKMNEKLYTVCDVAMNITMSKSTYVLSLEPKDPLVPARFTQ0PKN	480
Qy	1105	FSNKNKNTLPPEMKSFPTGPKRTINVLGAVKKPLSSAKOSOTKSSRRETVSNMASSNSP	1166
Db	481	FSNKNKNTLPPEMKSFPTGPKRTAVNLGAVKKPLSSAKOSOTKSSRRETVSNMASSNSP	540
Qy	1165	SSPGRIGRLDSSSEMDHSENDYMSPLPKKSDUKRDSLPULVRSLELKPGRKKTPTYE	1224
Db	541	SSPGRIGRLDSSSEMDHSENDYMSPLPKKSDUKRDSLPULVRSLELKPGRKKAPTYD	600
Qy	1225	QEBKLGMDLTKLVQEOKPKGSORSKRGHATASEDE0QWPEBKRLKEDILLENED0NSP	1288
Db	601	PEBKLGMDLTKLVQEOKPKGSORSKRGHATASDE0QWPEBKRNKKEELLENEDE0NSP	660
Qy	1285	PKKGRGRRPPLVGGTPEKEPTMTKSKGSKKSGPARPEEEEEE0SGTTEOKSSK	1344
Db	661	PKKGRGRRPPLVGGTSEKEPTMTKSKGKKKLVPPVVDDEEE0NGTTEHKSNSK	720
Qy	1345	QHVRYSRAQ0PAESPESALISTOSTPOKGGKGR0SKTSPS0PKKNV	1391
Db	721	QHVRYSRAQ0PAESPETSAVESTOSTPOKGGRGRESKASPS0PKKI	767

RESULT 6	ID	PRELIMINARY;	PRT;	701 AA.
08BJ18	08BJ18			
AC	08BJ18:			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DE	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	49J10.1.2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Thymus;			
RX	MEDLINE=22354683; PubMed=12466851;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs."			
RL	Nature 420:563-573(2002).			
DR	EMBL: AK041682; BAC31031.1; "			
SO	SEQUENCE 701 AA; 78714 MW; FAF6F0351AF7869B CRC64;			
Query Match	48.0%.	Score 3456;	DB 11;	Length 701;
Best Local Similarity	98.8%;	Pred. No. 1.2e-191;		
Matches 671;	Conservative 1;	Mismatches 7;	Indels 0;	Gaps 0;
QY	530	MYTRRNLPDPRGKADDFMKKFFQVLEDEDEKIRKOLEVLVSPSCSKQACGCVREITKKIGN	589	
DB	1	MYTRRNLPDPRGKADDFMKKFFQVLEDEDEKIRKOLEVLVSPSCSKQACGCVREITKKIGN	60	
QY	590	PKOPTNPFLEIKRFLERIPAVHIDTESISALIKOVNKSIDGTDADDEGVPPTOAIRAG	649	
DB	61	PKOPTNPFLEIKRFLERIPAVHIDTESISALIKOVNKSIDGTDADDEGVPPTOAIRAG	120	
QY	650	LELLKVLSTFTHPIISFHSAETTESILACIKMDEKVAEAAQIIFNTGSKIEDEPHIRSA	709	

RP SEQUENCE FROM N.A.
RC Tissue-Testis;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC039236; AAH39236.1; -
SQ SEQUENCE 529 AA; 61414 MM; 5C50973ACB920725 CRC64;

Query Match 35.0%; Score 2517; DB 4; Length 529;
Best Local Similarity 98.0%; Pred. No. 1.6e-137;
Matches 488; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

```
QY 1 MAHSKTRTNDGKITPPGKVEKISDKISKEMVRRLKMYVTQMDDODSEEEKELYLMA 60
DB 1 MAHSKTRTNDGKITPPGKVEKISDKISKEMVRRLKMYVTQMDDODSEEEKELYLMA 60
QY 61 LHLASDFLKHGKDVRLVAVACCLADIFRIYAPAPYTSPOKLKDIEMFITROLKGLDPT 120
DB 61 LHLASDFLKHGKDVRLVAVACCLADIFRIYAPAPYTSPOKLKDIEMFITROLKGLDPT 120
QY 121 KSPQENRYFYLENIAMVKSYNICFELDSNEIFOLYRTLESVINNGHOKVHMVMDL 180
DB 121 KSPQENRYFYLENIAMVKSYNICFELDSNEIFOLYRTLESVINNGHOKVHMVMDL 180
QY 181 MSIIICEDVYSOELDTVLVNLVPAHKNLKOAYDLAKALKRTAQALIEPIYTFEYNOV 240
DB 181 MSIIICEDVYSOELDTVLVNLVPAHKNLKOAYDLAKALKRTAQALIEPIYTFEYNOV 240
QY 241 LMLCKTSSIDSEHFVLDLILYLNIDSHLLSVLPQLEFKLSNDNEERLOVVKLLAMF 300
DB 241 LMLCKTSSIDSEHFVLDLILYLNIDSHLLSVLPQLEFKLSNDNEERLOVVKLLAMF 300
QY 301 GAKOSELASONKPLMOCTLGRFNIDHPIRLECYKFAHSCMLNHPDLAKDLTEYLKXVSH 360
DB 301 GAKOSELASONKPLMOCTLGRFNIDHPIRLECYKFAHSCMLNHPDLAKDLTEYLKXVSH 360
QY 361 DPEAIRHDVIVSYITAAKDDILVNDHLNFRVERTDCKRMRVREKAMGLAQYKKYA 420
DB 361 DPEAIRHDVIVSYITAAKDDILVNDHLNFRVERTDCKRMRVREKAMGLAQYKKYA 420
QY 421 LOSAAGKDAKOIAMIKDKLHIIYONSIDRLVERIFPAQYVPHNLETERMKCLYLL 480
DB 421 LOSAAGKDAKOIAMIKDKLHIIYONSIDRLVERIFPAQYVPHNLETERMKCLYLL 480
QY 481 YATLDLNAVKALNEMMKC 498
DB 481 YATLDLNAVKALNEMMKC 498
```

RESULT 9

```
QY 096KV4 PRELIMINARY; PRT; 479 AA.
AC 096KV4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 49310.1.2 (Androgen-induced prostate proliferative shutoff associated
  protein, isoform 2) (Fragment).
CN 49310.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z84572; CAC94787.1; -.
FT NON-TER 1
SQ SEQUENCE 479 AA; 54103 MM; AB57E74AAE9656DF CRC64;
```

Query Match 34.6%; Score 2487; DB 4; Length 479;
Best Local Similarity 100.0%; Pred. No. 7.7e-136;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 913 DECVOYROVFAOKLHKGISRLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 972
DB 1 DECVOYROVFAOKLHKGISRLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 60
QY 973 LKOHAAVSEKLSLPLEYVVPYTHLHAHPDYKVVQDIEOLKDVKECLMFWLEITLMAKN 1032
DB 61 LKOHAAVSEKLSLPLEYVVPYTHLHAHPDYKVVQDIEOLKDVKECLMFWLEITLMAKN 120
QY 1033 ENNSHAFIRKVENIKOTKDAQGPDDAKMNEKLYTVCVANNIIMSKSTYVSLESKDPV 1092
DB 121 ENNSHAFIRKVENIKOTKDAQGPDDAKMNEKLYTVCVANNIIMSKSTYVSLESKDPV 180
QY 1093 LPARFTQPDKNFSNTKTYLRLPEPKMSFTTPCKPRTTNVLGAVNKPPLSSAGOSQTKSSRM 1152
DB 181 LPARFTQPDKNFSNTKTYLRLPEPKMSFTTPCKPRTTNVLGAVNKPPLSSAGOSQTKSSRM 240
QY 1153 ETVSNASSSSNPSSPGRIKGRDSEMDHSEMDYTMSSPLPGKKSDKRDSDLYRSELE 1212
DB 241 ETVSNASSSSNPSSPGRIKGRDSEMDHSEMDYTMSSPLPGKKSDKRDSDLYRSELE 300
QY 1213 KPRGRKKTPTVTEOEKLGMDLTKLVQEQKSGORSRKRGHHTASESDQOMPEBKRLKE 1272
DB 301 KPRGRKKTPTVTEOEKLGMDLTKLVQEQKSGORSRKRGHHTASESDQOMPEBKRLKE 360
QY 1273 DILENEDQNSPPKKGKGRPPKPLGGCTPKPEEPTPMKSKGSKKSGPPAPEEEREER 1332
DB 361 DILENEDQNSPPKKGKGRPPKPLGGCTPKPEEPTPMKSKGSKKSGPPAPEEEREER 420
QY 1333 QSGNTEOKSKSKOHVSRRAQRAESPSSAIESTOSTPQGRGRSTPSPQPKKRV 1391
DB 421 QSGNTEOKSKSKOHVSRRAQRAESPSSAIESTOSTPQGRGRSTPSPQPKKRV 479
```

RESULT 10

```
QY 096KV3 PRELIMINARY; PRT; 536 AA.
AC 096KV3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 49310.1.1 (KIAA0979, isoform 1) (Fragment).
CN 49310.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z84572; CAC94788.1; -.
FT NON-TER 1
SQ SEQUENCE 536 AA; 60577 MM; 3A30B33F56C81B8C CRC64;
```

Query Match 34.4%; Score 2476.5; DB 4; Length 536;
Best Local Similarity 99.8%; Pred. No. 3.6e-135;
Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 913 DECVOYROVFAOKLHKGISRLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 972
DB 1 DECVOYROVFAOKLHKGISRLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 60
QY 973 LKOHAAVSEKLSLPLEYVVPYTHLHAHPDYKVVQDIEOLKDVKECLMFWLEITLMAKN 1032
DB 61 LKOHAAVSEKLSLPLEYVVPYTHLHAHPDYKVVQDIEOLKDVKECLMFWLEITLMAKN 120
QY 1033 ENNSHAFIRKVENIKOTKDAQGPDDAKMNEKLYTVCVANNIIMSKSTYVSLESKDPV 1092
DB 121 ENNSHAFIRKVENIKOTKDAQGPDDAKMNEKLYTVCVANNIIMSKSTYVSLESKDPV 180
QY 1093 LPARFTQPDKNFSNTKTYLRLPEPKMSFTTPCKPRTTNVLGAVNKPPLSSAGOSQTKSSRM 1152
DB 181 LPARFTQPDKNFSNTKTYLRLPEPKMSFTTPCKPRTTNVLGAVNKPPLSSAGOSQTKSSRM 240
QY 1153 ETVSNASSSSNPSSPGRIKGRDSEMDHSEMDYTMSSPLPGKKSDKRDSDLYRSELE 1212
DB 241 ETVSNASSSSNPSSPGRIKGRDSEMDHSEMDYTMSSPLPGKKSDKRDSDLYRSELE 300
QY 1213 KPRGRKKTPTVTEOEKLGMDLTKLVQEQKSGORSRKRGHHTASESDQOMPEBKRLKE 1272
DB 301 KPRGRKKTPTVTEOEKLGMDLTKLVQEQKSGORSRKRGHHTASESDQOMPEBKRLKE 360
QY 1273 DILENEDQNSPPKKGKGRPPKPLGGCTPKPEEPTPMKSKGSKKSGPPAPEEEREER 1332
DB 361 DILENEDQNSPPKKGKGRPPKPLGGCTPKPEEPTPMKSKGSKKSGPPAPEEEREER 420
QY 1333 QSGNTEOKSKSKOHVSRRAQRAESPSSAIESTOSTPQGRGRSTPSPQPKKRV 1391
DB 421 QSGNTEOKSKSKOHVSRRAQRAESPSSAIESTOSTPQGRGRSTPSPQPKKRV 479
```

```
Db 181 LPARFTQDPDKNSFNKNLYLPPMKSFPTPGKPKRTTNVLGAVNKKPLSSAGKQSKSSRM 240
QY 1153 ETVSNASSSSNPSPGRIKGRLDSEMDHSENEDEYTMSSPLPGKSKSDSDVRSLELE 1212
Db 241 ETVSNASSSSNPSPGRIKGRLDSEMDHSENEDEYTMSSPLPGKSKSDSDVRSLELE 300
QY 1213 KPRGRKTPVTEDEBKIGMDLTKLYOEQPKGSQSRKRGHTAASEDEQWPEEKRLKE 1272
Db 301 KPRGRKTPVTEDEBKIGMDLTKLYOEQPKGSQSRKRGHTAASEDEQWPEEKRLKE 360
QY 1273 DIENEDENSPPKKGRGRRPPKLGSTPKPEPTMTSKGSKSGSPAPPEEEER 1332
Db 361 DIENEDENSPPKKGRGRRPPKLGSTPKPEPTMTSKGSKSGSPAPPEEEER 420
QY 1333 QSGNTEOKSKSKOHVRSRAAO--RAESPSSAISTESTSTPOKGRGRPKTSPSPQPKNV 1391
Db 421 QSGNTEOKSKSKOHVRSRAAOQSRAESPSSAISTESTSTPOKGRGRPKTSPSPQPKNV 480

RESULT 11
Q8N7J4 PRELIMINARY; PRT; 600 AA.
ID 08N7J4:
AC 08N7J4:
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ41012.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=uterus;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Saito H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098331; BAC05286.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 600 AA; 69000 MW; 12917E7912C86BFD CRC64;

Query Match 34.4%; Score 2474; DB 4; Length 600;
Best Local Similarity 80.6%; Pred. No. 5.9e-135;
Matches 460; Conservative 64; Mismatches 47; Indels 0; Gaps 0;

QY 10 DGIITPPGVKEISDKISKEWVRRLKMYKTFMDMDQDSSEKELYLNALHLASDFL 69
Db 20 DGIITPPGVKEISDKISKEWVRRLKMYKTFMDMDQDSSEKELYLNALHLASDFL 79
QY 70 KHQEKDVRLLVACCLADIFRIYAEADYTSIDKIDIFLFTTRQKLEDTKSPQFNRYE 129
Db 80 RNPKDVRLVACCLADIFRIYAEADYTSIDKIDIFLFTTRQKLEDTKSPQFNRYE 139
QY 130 YLLENIAWKSYNICFELEDSENEIFQTLYRTLFSVINNGHOKYHMHVMDLMSITGEGD 189
Db 140 YLLENIAWKSYNICFELEDSENEIFQTLYRTLFSVINSHKKQOMHMLDMSITGEGD 199
QY 190 TVSOELDLTVLVNLVPAHKNLNKOAYDLAKALKRQAQATEPYTYTFENQVLMGKTSIS 249
Db 200 GAVQELDLSTILINIPAHKNLNKOSFDLAKVLLKRYQTITACIANFNQVLYGRSSVS 259
QY 250 DLSEHVEDLLELYNIDSHLLSVLPQLEFKLSNDNEERLQVYKLLAKMGAKDSGLAS 309
Db 260 DLSEHVEDLLELYNIDSHLLSVLPQLEFKLSNDNEERLQVYKLLAKMGAKDSGLAS 319
QY 310 QNRPVMOCYLGRFNDIVNFRLECYVRFASHCLMNHPRDLAKDLYLVKRSHPDEALRHD 369
Db 320 QNRPVMOCYLGRFNDIVNFRLECYVRFASHCLMNHPRDLAKDLYLVKRSHPDEALRHD 379
```

```
QY 370 VIVSIYTAARKDILLVNDHLNFEVERTLDKRRVRREKAMMGIAQIYKRYALOSAGKDA 429
Db 380 VIVSIYTAARKDILLVNDHLNFEVERTLDKRRVRREKAMMGIAQIYKRYALOSAGKDA 439
QY 430 AKQIAWTKDKLLHYTONSIDDLVRYEITPAQYVWPHNLETTERMKCLYLYATLDLNAV 489
Db 440 AERYSWIKDKLLHYTONSIDDLVRYEITPAQYVWPHNLETTERMKCLYLYATLDLNAV 499
QY 490 KALNEMKCONLRHGVKDDLDIKOPTASVKAIPSKVWVITRNLPDGAQDPFAKRF 549
Db 500 KALNEMKCONLRHGVKDDLDIKOPTASVKAIPSKVWVITRNLPDGAQDPFAKRF 559
QY 550 TOVLEDEKIRKQLEVLVSPSCQKQAGCV 580
Db 560 NOVLDGDEKLRQLELLISPTCSCKQADICV 590

RESULT 12
Q8VDS0 PRELIMINARY; PRT; 584 AA.
ID 08VDS0:
AC 08VDS0:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Hypothetical 65.8 kDa protein.
GN 9030416H16RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021408; AAH21408.1; -.
DR MGI; MGI:1918771; 9030416H16RIK.
KW Hypothetical protein.
SQ
SEQUENCE 584 AA; 65801 MW; C7015B0681432851 CRC64;

Query Match 32.2%; Score 2317; DB 11; Length 584;
Best Local Similarity 74.3%; Pred. No. 6.8e-126;
Matches 434; Conservative 74; Mismatches 76; Indels 0; Gaps 0;

QY 525 IFSKVAVITRNLPDGAQDPFAKRFQVLEDEDEKIRKQLEVLVSPSCQKQAGCV 584
Db 1 MFGKLMITANLPDGAQDPFAKRFQVLEDEDEKIRKQLEVLVSPSCQKQAGCV 60
QY 585 KKLGNPKQPTNPFLEMIKFLERIPVHIDTESISALIKOVNKSIDGTADDEGVPTDO 644
Db 61 KKLGNPKQPTNPFLEMIKFLERIPVHIDTESISALIKOVNKSIDGTADDEGVPTDO 654
QY 645 AIRAGLELLVSTTHSIFSHSAFTFESLACLKMDKEKAAEALQIFKMGSKTEEDFP 704
Db 121 AIRAGLELLVSTTHSIFSHSAFTFESLACLKMDKEKAAEALQIFKMGSKTEEDFP 714
QY 705 HIRFALLPVLNHKKKPPROAKYVAGIHAIFSSCTOPAQIPEPLHKSIDSNLEHLI 764
Db 181 QIRSTLPIHLQAKKRGTPHQAKQAVCHAHAFENKVEQLAQIFEPILSRSLNDVPRQLI 240
QY 181 QIRSTLPIHLQAKKRGTPHQAKQAVCHAHAFENKVEQLAQIFEPILSRSLNDVPRQLI 240
QY 765 TPVLTIGHIALPADQFAAPKSWVAFTYVKDLLMNDRLPGKTKTKLVNDEEVSPEPMV 824
Db 241 TPVLTIGHIALPADQFAAPKSWVAFTYVKDLLMNDRLPGKTKTKLVNDEEVSPEPMV 834
QY 825 KIQAIKMYVRLGKMKNNHSSGSTYRLTLTIHSGDITDEGKISKPMRSRLRLAGS 884
Db 301 KQVAIKILVRLGKMKNNHSSGSTYRLTLTIHSGDITDEGKISKPMRSRLRLAGS 894
QY 885 AIYVLAQPCYHEITTEYOQLCALINDEQYOVQYFAOKLHGLSRLRLPLEYMAICA 944
Db 361 AIYVLAQPCYHEITTEYOQLCALINDEQYOVQYFAOKLHGLSRLRLPLEYMAICA 954
QY 945 LCAKDPVKERRAHARQCLVKNINVRREYLRKQHAHVSEKLLSLPEYVVVPTIHLADPD 1004
```

```

|||||
421 LCAPKPKERHAHQCLIKNTSIRREYIKONPMATEKLSLPEVVPYMHHLAHPD 480
QY 1005 YKVDIOELQDKVCKCLFVLEITLMAKNENNSHAFIRKVENIKOTKDAQGPDKAKMNEK 1064
DB 481 FTRSDVQDLQRIKCECLMFEVLEMTKKENNSHAFMKWAENIKILTRAQSDPDESKTNEK 540
QY 1065 LTYVCDVAMNIIMSKSTYTSLESKPDVLPAPFFQOPKNSFT 1108
DB 541 LTYVCDVALCVINSKSAICNADSPKDPVLPMPKFFQPEKVISLT 584

RESULT 13
09V646 PRELIMINARY: PRT: 1218 AA.
AC 09V646:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG17509 protein.
CN CG17509.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houtlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodman T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003823; AAF58593.1;
DR FlyBase: FBgn0033676; CG17509.
SO SEQUENCE 1218 AA; 138887 MW; F4AGB7238A0801C2 CRC64;

```

Query Match

27.6%; Score 1982.5; DB 5; Length 1218;

```

Best Local Similarity 37.3%; Pred. No. 4e-106;
Matches 444; Conservative 247; Mismatches 421; Indels 77; Gaps 22;
QY 13 ITPPGVKEISDKISKEKEMVRLLKVVTKFMDQDSEEEKELYNLALHLASDFELKHP 72
DB 4 IYPTGCRPLVEDLCTDELIRRLKTLANVLAQTMDDQDNLVQO-YIPLAHLHLDDEFMQHP 62
QY 73 GKDVLVAACCLADFRIPAPAPYTSYDPLKLMFMTITROLKGLGEDTRKSPQPNVFTYL 132
DB 63 SRDVOILLACCAVDLRYAPAPAPYKEDQITIKFELKQLHGLKDPDPSEFKRYFYL 122
QY 133 ENIAWKSYNICEFELEDSNEIFOTLYRTLFYSVINGNHNOKVMMHNVLDSSITGDFVS 192
DB 123 ENIAWKSYNICEFELEDSNEIFOTLYRTLFYSVINGNHNOKVMMHNVLDSSITGDFVS 182
QY 193 QELDVTLVNIVPAHKNLNKQAYDLAKALKLRTQAQIEPYITTFEVOYLMGK--TSSD 250
DB 183 VELLDLILINIVEPKSNKKFACQLTQELITFTGDALSTIKMFENRALVMDKPMTKLS- 241
QY 251 LSEHFEDLILEYNDSHLLSVLPQLEFKLSNNEERLQVYKLAKMFGAKDSLSAQ 310
DB 242 ITNKIYDIYIELNRIAGLCSVLPQLEFKLSNNEERLQVYKLAKMFGAKDSLSAQ 301
QY 311 NKPLQCYLGRFNDIHVPIRECVKFAASHCLMHPDLAKDLTEYLKVSHPDEEAFIRHDV 370
DB 362 VVAIVETAKRDEFTLEAPDLLEIVERTLDKKYKIRRDAMGLAYIKRAICEPNDIST 421
QY 429 AAK-QIAMIKOKLHITTYONSIDDLVYERIFAOYVNHNETTERMKCYLYLVATDLN 487
DB 422 GLKVVADIKNKLHLYGVYKGLDELRLLITGLVPAKLAPEEEMKLVLLDLADLAN 481
QY 488 AVKALNEMKCONLRHOKVLDLIDIKOPK--TDAVSKAIFSVVYITRNLPPGKAODFM 546
DB 482 ATKAVLELOKQNMKTRNVSDMKIKLHNSKEFTPRVLSQSAKQANIAKLDPRLAAEYL 541
QY 547 KKEFVLEDEDEKIRKOLELVVSPSCQAOBSCVBEITKKLGPNKPPNPLEMIKFLLE 606
DB 542 TQFSNNLRKDKOLLICINILVKRDSRECADTQVGLLKIGAHVQ-SNLVYNTVKMIE 600
QY 607 RIAPPHIDPESIALIKOVNSIDGTADDEDEGVTTDAIRAGLELLAVLSTPHISHS 666
DB 601 RYASVAVKESIGVLSLIEQCTEKGSCEETIGISAQEGERGKLLMLLSVESAHPFT 660
QY 667 AETFEISLACIKMDEKVAEALQIFKNTG--SKIEEDPHIRSAIPLVHLNKKSGPP 723
DB 661 DLSLRLSLISLSEEDYVAPLVKLTILHGRQPLIDPPTAIDELAPVCKDFALICTP 720
QY 724 ROAKYAINCIH-----AIFSSKEIO-----FAQIEPLKSLDPSNLEHL 763
DB 721 KOAKHAYACIPFNSOSSASTGATGASASASTGTQVHPINIEIETRLKLTLP-NCEHO 779
QY 764 ITPVLTGHIHLLAPDGAAPAKSMVAFIYKDLMLNRLBGGKKTTL-----VWPDDEV 818
DB 780 RKTITLGHIAVNMQALFLTPKNNIARIYKELLIOE-VPAQRYEYELPEDSDWQAQKYL 838
QY 819 SPETVVKIQAIKMMVRMLGKMKNNHSGSTSLRLITLISDGDITRQGISKPDMSRL 878
DB 839 PPDLTKCKDALKAMRWLGLTRTD-EHAAQKTFRLAFAVNOGRGLDQONRLCGAKRWL 897
QY 879 RLAAAGSAIVKLAQEPCEYHITTELOYOLCALAINDECYOVNOVEAQKLHKGLSLRL- 934
DB 898 RLGAACAMKLVKCEQGVGDQYSAEQYLOLSOLMDPVEVEIIFARKLHKGLSLRPNRC 957
QY 935 LPLEYMATICALCAODPVKERRAHNQCLVKNINVRREVLKQHAANS-----EKLSLTP 988
DB 958 LPLDFMGLVYVLAGLETERRKLDLVNHYAETVNNKREELKTVAMTSPDSSSTESOSLHLP 1017
QY 989 EYVVPYTIHLAHPDQYKVDIEQLDKVCKCLFVLEITLMAKNENNSHAFIRKVENIK 1048

```

```

Db 1018 DYMLAFPIPLVHDPRFTNEDVYVQLRMEKCLFLEPLMAKRETVHSHYKOLLQILK 1077
QY 1049 QTRDAQDPDDAKNNEKLYTVCDAVAMNIMSK-----STYLSLESPKDPVLPARFPTOP 1101
Db 1078 HREBSLG-SPKRDNYKMMALCDLALMYLIDSKFSPDGNSTFSMPL-----ALPEMYKKEP 1132
QY 1102 D-KNFSNTKRYLPPKMSFTPKPKTTNVLGAVNKLSSAGKOSQTKS 1149
Db 1133 AVANFONNDVYIPLDVT-----LGA--KSTSKAAVAMTMS 1167

RESULT 14
Q9HFF5 PRELIMINARY; PRT; 1205 AA.
ID Q9HFF5;
AC Q9HFF5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Protein with possible role in chromosome dynamics by similarity to
DE yeast pds5.
OS SPAC110.02 OR PDS5+.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972n-;
RA Zimmerman W., Wandt R., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LG72;
RC MEDLINE=2148173; PubMed=11598020;
RA Tanaka K., Hao Z., Kai M., Okayama H.;
RT "Establishment and maintenance of sister chromatid cohesion in fission
RT yeast by a unique mechanism.";
RL EMO J. 20:5779-5790(2001).
DR EMBL; AL441624; CAC08560.1; -.
DR EMBL; AB067651; BAB1784.1; -.
DR GeneDB; Spombe; SPAC110.02; -.
SQ SEQUENCE 1205 AA; 138874 MW; 935DABDE0A5E30FF CRC64;

Query Match 11.6%; Score 833.5; DB 3; Length 1205;
Best Local Similarity 23.3%; Pred. No. 1.1e-39;
Matches 296; Conservative 264; Mismatches 564; Indels 149; Gaps 38;

QY 26 ISKEWVRRLKMYVKTMDMDQDSEKELYLNLALHLASDFLKHFGKDYRLVACCLA 85
Db 18 LTRSEILKRLRDLGELTSLSDPT-IDRDSVLPVARSLLVNNMLLHDKGIRSYTLCICV 76
QY 86 DIPRIYAEAPYTSPPDKIDIFMFTROLKGLDEDTKPOFNRYRYLLENTIAMVSYNICE 145
Db 77 ELRLCPADAPFTL-SQLEDFIYVITLKILSGIMNOBSTYTPQIYELIESNNVSAVLIY 135
QY 146 ELEDSENIETQLYRTLESVINNGHOKVHHMVDMLSSICEGDTVSOELIDYLVNLV- 204
Db 136 DLPAEELVNIIFRLFDLARKGTNKVEFYMLDIINQLINEITIPAAANLILFADLIS 195
QY 205 -----PAKRLNKQATDLAKALKKRTQALETPTTFEFOVLMGKTISD---LS 252
Db 196 GKGVROTIGSSDSTNHGPAFLAKNIPHDSADRLORVCOYFSDIIFDSRSLSDSMTP 255
QY 253 EHVE--DLILELYNIDSHLLSVLPQLEFKLKSNDNEBRLOVYLLAKMGAKDSELSQ 310
Db 256 EPIFSNHLVQLMKYATTLTINIIPQENELQAGTQSVRLVALETGMLL--QNALMSD 313
QY 311 NKPLMOCYLGRIRLIRLECYKFAASHCLMNH--PLDAKLTLEYLKVASHDPEAIR 367
Db 314 YPRVMSAFGCRLNDKSVACRIKTEVANSMLQNSLANSSEIENNVVQALQSLADPTDKVR 373
QY 368 HDVIVSLVTAAKADI-LLVNDHLNLFVREKRLDKRMVRYKREAMGALQIYKKTALQSAAG 426

```

```

Db 374 VAILKTEIQLTEFEFKQFVSQVALKIMGDLRBRKLVNRQALITLSQIYNRAVQDLIDG 433
QY 427 KDAKQI-AMIKDKLILHYVONSIDRLVERIFPAQVNH-NLETERKCCYIYATL 484
Db 434 VESIQWFSNIPSSLEVFYVNDETTAAVEICMAELVLOYLSSDITRLNRLFLSLIKYF 493
QY 485 DLNAVKA-----LNMKCOMLLRHQV-----DLIDLKQPKTASVAIFSKVAVIT 533
Db 494 SEKAMRVFILLQOVYSELLNYIIBCKKNYNGVANDNDEESITNKKVYI---DLIS 549
QY 534 RNLPDCKADQPMKFTQVLEDEKIRKQLEVLYSPSCSKQAEQCVREITTKLGNKOP 593
Db 550 SKSSNPILTATPFKFAE-LNDROSYKMLQ-----PFSIKSEYQVAV----- 590
QY 594 TNPLFMKIFLEIRIAVNHIDTESI-----SALIKOVNSIDTADDEDEGVPTDQ 644
Db 591 -----LKSITKILFKVSET-LSIASLECFRIFVRSALFARNKSNVHEITQLNEPVAYH 645
QY 645 AIRAGLELKVLSFTPHDISFHS--ETTFESLACLKMDKDEVAEALQIFKNTGSKLEED 702
Db 646 FLKSEALDQHLPLIHNINIGEYVIEVENILVSSIGESDPKVIKALSQF-----SKRKN 700
QY 703 FPHIRSLALPYLHNKSKKPPROAKYAIHCIIHAIFSSKEPQAFQIEPLKSLDPSNLEH 762
Db 701 F-SIQTTAELIRKLCHGTQEQAKQAATII-ATTEKREKLDITNIV-----ENLEY 752
QY 763 ---LTPPLYTIGHALLAPDOFAPMKSWATFVKDILLMNDRLPKG-----KTKKLWPD 815
Db 753 NGGILPVLMILGQLEFLLTEE-VEKAVDQTEFLVKVYI--QRPPEKYDTHNDEMCY 809
QY 816 EEVSPETMVXIOAIKMYVRLMGKNHNS--KSGSTRLRLTLTLHSDGLTEGKISKP 873
Db 810 EKLNLTMCKVLAIRVLNRLRAAGTEALNIGAPIIKILKVLVLMADGELSPKPKPI 869
QY 874 DMSRLRAAGSAIYKLAQEPCHHEITLLEQYQCALAINDECVYRQVPAKILKIGLSRL 933
Db 870 SRAYLRITASKYFLKLSIFPYAEHIDFSSVQISLCOBENDVRLMFLTKLOKOLQ 929
QY 934 RLPLEYMAICALCKADPVKERRARAGCLVKININREYIKONAASVEKLLSLPEYVVP 993
Db 930 KLPISTYPLFLTAIVDEEELIKTA-----SIWRISOVAFFQKHDFMEVAVT 978
QY 994 YTHLHADDDYVQDIEQDKVCEIWPVLEITLMAKNENNSHAFIRKVENIKQTKDA 1053
Db 979 YLHLLSHHPDISSIESENSIDFAYIRFYVDVNVSENV---PIYFHLMRIKQSYDV 1034
QY 1054 QGPDQAKNNEKLYVVCYVAMNIMSKS-----TYY--SLESKDPVLPARFPTOPDKN 1104
Db 1035 --IEDG--NNYIYVLSMAOKILQVKSQNGWSLTYPKQIKLPEYELIRPISIDEKKRI 1090
QY 1105 FSNTKNVLPPKMSFTPKPKTTNVLGAVNKLSSAGKOSQTSFSEMETVSNASSSS 1164
Db 1091 FN-KITITTKMES-----QIEHAIRTPVPSAKOTTKNHAUKOKKTHSSKSDK 1138
QY 1165 SSPGRIGRLDSEMDHSENEEDYTMSSPLPGKSKDKRDSDLVASELEKPRGRKTPYTE 1224
Db 1139 KSSRRK-----NEKRRKLNQNPININVERBSSRFGAIRINSEAPS-----SSEISE 1189
QY 1225 QEEKLGDDDLTKL 1237
Db 1190 EEEIESEDEDEI 1202

RESULT 15
Q9FILO PRELIMINARY; PRT; 1638 AA.
ID Q9FILO;
AC Q9FILO;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Gb|AA07790.1.
OS Arabidopsis thaliana (Mouse-ear cress).

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 16:31:58 : Search time 20 Seconds

(without alignments)
3270.707 Million cell updates/sec

Title: US-09-512-581b-2

Perfect score: 7193

Sequence: 1 MAHSKTRNDKIRTPGVK.....QKGRGRSKTPSPQPKKNV 1391

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	833.5	11.6	1205	1	PD55_SCHPO
2	557	7.7	1277	1	PD55_YEAST
3	239.5	3.3	5171	1	BPEA_HUMAN
4	235.5	3.3	1875	1	MLP1_YEAST
5	229	3.2	2663	1	CENE_HUMAN
6	227	3.2	1916	1	RIF1_YEAST
7	220.5	3.1	1727	1	ALM1_SCHPO
8	217.5	3.0	1020	1	NFR_HUMAN
9	217	3.0	1790	1	USO1_YEAST
10	212.5	3.0	8545	1	ANCL1_CAEEL
11	212	2.9	831	1	NFR_RAT
12	210.5	2.9	5586	1	MDN1_HUMAN
13	209	2.9	7389	1	BPA1_MOUSE
14	208	2.9	1153	1	A3D1_HUMAN
15	206.5	2.9	699	1	NPL4_HUMAN
16	206	2.9	6632	1	UN89_CAEEL
17	205.5	2.9	2230	1	GOG4_HUMAN
18	205	2.8	2349	1	TBR_HUMAN
19	203.5	2.8	2459	1	MABP_RAT
20	202	2.8	2748	1	NDM1_YEAST
21	201	2.8	700	1	TRDN_CANFA
22	199	2.8	1359	1	ATRX_CAEEL
23	199	2.8	2464	1	MABP_MOUSE
24	197.5	2.7	3259	1	GIAN_HUMAN
25	195.5	2.7	1679	1	Y109_YEAST
26	195	2.7	488	1	CYL2_BOVIN
27	194	2.7	705	1	TRDN_RABIT
28	194	2.7	1533	1	TP2A_PIG
29	194	2.7	1539	1	Y373_HUMAN
30	193.5	2.7	3210	1	CENF_HUMAN
31	192.5	2.7	2469	1	TEB0_HSVSA
32	190.5	2.6	2468	1	MABP_HUMAN
33	188.5	2.6	633	1	MLH1_TETPH

ALIGNMENTS

RESULT 1	ID	PD55_SCHPO	STANDARD:	PRT: 1205 AA.	
AC	Q9HFE5	O94237;			
DT	15-SEP-2003	(Rel. 42, Created)			
DT	15-SEP-2003	(Rel. 42, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	Sister chromatid cohesion protein pds5 (Precocious dissociation of sisters protein 5).				
DE	Sisters protein 5).				
GN	PD55 OR SPAC110.02.				
OS	Schizosaccharomyces pombe (fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OX	Schizosaccharomycetes.				
NCBI_TaxID=4896;					
[1]					
SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.					
RC	STRAIN-972;				
RX	MEDLINE=21481773; PubMed=11598020;				
RT	Tanaka K., Hao Z., Kai M., Okayama H.;				
RT	"Establishment and maintenance of sister chromatid cohesion in fission yeast by a unique mechanism.";				
RL	Embo J. 20:5779-5790(2001).				
[2]					
SEQUENCE FROM N.A.					
RC	MEDLINE=21848401; PubMed=11859360;				
RX	STRAIN-972;				
RA	Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Sgouras K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,				
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,				
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,				
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jags K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,				
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,				
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,				
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,				
RA	Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,				
RA	Woodward J., Volkert G., Aert R., Robben J., Grymoprez B.,				
RA	Wolters I., Vansirels E., Rieger M., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,				
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,				
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Punnelle B.,				
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,				
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,				
RA	Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,				
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,				
RA	Domínguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,				
RA	Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,				
RA	Shpakowski G., Ussery D., Barrett B.G., Nurse P.;				
RT	"The genome sequence of Schizosaccharomyces pombe.";				
RL	Nature 415:871-880(2002).				
[3]					
SEQUENCE OF 816-1205 FROM N.A.					
RP	Lee M., Yoo H.S., Chung K.S.;				

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP FUNCTION.
 RX MEDLINE-21850851; PubMed-11861765;
 RA Mang S.-W., Read R.L., Norbury C.J.;
 RT "Fission yeast Pds5 is required for accurate chromosome segregation
 and for survival after DNA damage or metaphase arrest.";
 RL J. Cell Sci. 115:587-598(2002).
 CC -1- FUNCTION: Required for the establishment and maintenance of sister
 chromatid cohesion during S phase. Prevents their formation until
 esol is present. May also have a role during meiosis.
 CC -1- SUBUNIT: Interacts with esol.
 CC -1- SUBCELLULAR LOCATION: Nuclear; centromeric. Localized to chromatin
 throughout the cell cycle.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL: AB067651; BAB71784.1; -
 DR EMBL: AL441624; CAC08560.1; -
 DR EMBL: AF049529; AAD02493.1; -
 DR GeneDB_Spombae: SPAC110.02; -
 FM Mtosis; Nuclear protein; Centromere.
 FT CONFLICT 816 818 TMC -> GTS (IN REF. 3).
 FT CONFLICT 902 908 QISLCO -> SNLIMP (IN REF. 3).
 SQ SEQUENCE 1205 AA; 138874 MW; 935DABDE0A5E30FF CR664;
 Query Match 11.68; Score 833.5; DB 1; Length 1205;
 Best Local Similarity 23.38; Pred. No. 5.2e-31;
 Matches 296; Conservative 264; Mismatches 564; Indels 149; Gaps 38;
 QY 26 ISKEWVRILKMYKTFMDMDDOSEKELYLNLALSLASDFLKHGKYRLIVACCLA 85
 DB 18 LTTSSEILKRLDLGELTSLSDPT-IDRDSVLPARSLVNNNLHNDKGRSTYTCIV 76
 QY 86 DIFRIYAPARYTSPDKLIDFMETROLGLDPTKSPQNRVYLLLENIAWVSYNICEF 145
 DB 77 ELRLCAPDAPFTL-SQLEDIFQYILKILSLMNQESTYYPQIYELLESLSNVASAVLIY 135
 QY 146 ELEDSEITFQLYTLESVYNGHNGKQVHHMVDLMSITCEGDTVSOELDTVNLV 204
 DB 136 DLPVAEELVNIFFELFEDLARKGTTKNEVEYMDIINQINLETITPAALNIIFAOLIS 195
 QY 205 -----PAHKNLNQAYDLAKALKKPTAQATEPTTFPFNOVLKMGTSISD--LS 252
 DB 196 GKGVROITGSSDSINHGPAFOLARNIFHDSADRLQRYVCOTFSDITFDSRSLSDSMTTP 255
 QY 253 EHWV--DLLELYNIDSHLLSVLPOLPEFLKLSNDNERLOVYKLAKMGAKDSLASQ 310
 DB 256 EFLFSHNLVQLMKYAPFTLLIIPFENELQAOGTQSVRLVAITVGML--QONALWSD 313
 QY 311 NKPLMOCYLRPNDIHPRILECYKPSHCLMNH---PDLAKDLETLKYRSHDPEAIR 367
 DB 314 YPRWSAFCGRILDKRSVACRIKCEVANSALQNSLATSEIENVWMOQLSLADTDEKVR 373
 QY 368 HDVIVSVITAAKKDI--LLVNDHLLNFVREPTLDRMVRREKMGALQYKRYALQSAQ 426
 DB 374 VALTKTEQULTEFFKQFSVQALKMGDRLRDKLNVRLQALITTLISQITNRAVODLIDG 433
 QY 427 KDAAKOI-AMIKDKLHIYYQNSIDRLVERIFAQYVPH-NLEETTERKCLYLYATL 484
 DB 434 VESIQMFSWIPSLILEVFVYVNDETTNAAVEICAEVLVQLYSSDQTFLNRLFLSKYF 493
 QY 485 DLAAVKA---LDEMKGQNLRHQV-----DLIDLKQPKTASVKAITSKVAVIT 533
 DB 494 SEKAMRFILLLORQVYSELNLNYIECCNNGVNDNEESTTNKLKTYI---DIIS 549
 QY 534 RNLPDPEKADPFKKFQVLEDEKIRKQLEVLVSPSCSKQAGCVRREITTKKLGKNGKOP 593

DB 550 SKSSNPPLTEATYERKFAE-LNDROSYKMLQ-----TFSLKSEXOVV----- 590
 QY 594 TNPLEMIKFLTERIAVNIHIDTESI-----SALIKQVNSIIDTADDEGEVPTDQ 644
 DB 591 -----LKSITKILFKRVSET-LSTASLECFRIYVRSALFARNKSNVHITQLNPPVYHN 645
 QY 645 AIRAGLELLVLSFTHDISFHS--ETFESILACLKMDKDVAAEQALQIKNTGSKIEED 702
 DB 646 FLKPSALLQHLPLIHNINIGEVVIEVENIISGIESDPKVIKALSQF-----SKRKN 700
 QY 703 FPHTRSLDLLVNLHNSKGPGRQAKVAILHCIAHFSSKEFOFOIPEPLKSLDPSLKH 762
 DB 701 F-STQTTTAAELIRKLCLHGTEQEQAKQATII-ATTEFERKELMTITIV-----EMLEY 752
 QY 763 ---LITPLVTIGHIALIAPQOFAPRKSVAATFVKRLDLMNDRLPKG-----KTKKLVPD 815
 DB 753 NGGLPVLMLTLGQLFVLTLEE-VEKADVQTEFLYKVT--QRFPEKYDTHNDEKCTY 809
 QY 816 EEVSPETPVKIQAIKMYRWMLGKNHNS--KSGTSTRLTLTLHSDGLTDEGKISKP 873
 DB 810 EKLQNLTKCKVLAIRVLVNLRAAGTEALNIGAPITIKLKVLMADGELSFPKNTPKI 869
 QY 874 DMSRLLAAGSALYKLAQECYHEITILEYOUCALAINDECYQVROVFAQKILHKGISRL 933
 DB 870 SRAVLRILTAKRYFLKLSIPFYAEHIDFSSYQISLQDENDFVRLFLTKLOLOQLK 929
 QY 934 RLPLEYAICALCAKDPKERRARQCLVKININREYIKONAAVESEKLISLPEYVVP 993
 DB 930 KLPISTYPLFLFNAVDEDEELTKRA-----SIWRSQVAFQKTHDPTMEVAT 978
 QY 994 YTHLHAHPDYKQVQDIEOLKDYECLEFVLEITLAKNENSHAFIRKVAENIKQTKDA 1053
 DB 979 YLHLHLSHNDISSIESENSLDIFAYIRFYVTVVENSEN-----PIVFHLMORIKOSYDV 1034
 QY 1054 QGPDDAKMNEKLYVQCVAMNIIMSKS-----TV--SLESPKPDVLPAREFTOPDKN 1104
 DB 1035 --IEDG--NNITYVLSMAOKILOVKSQNGWSLTPPKOIKLPELRLPIPSIDEKRI 1090
 QY 1105 FSNTRKLYLPPEMKSFPTPGKPKTTNVLGAANKPLSSAGKOSQTKSSHMETVSNASSSNP 1164
 DB 1091 FN-KIITPKMES-----QIEHALRTVPSPFAKQTKTKHNAUKQKTHSSKSDK 1138
 QY 1165 SSPGRIRGRDSSMDHSENDITWSSPLDGKSKDKADDSDVLRSELEKRGKRTPYTE 1224
 DB 1139 KSSRRRK-----NEKRKRLQONPINRVBERSSRPQIGIRINSEAPS-----SSERISE 1189
 QY 1225 QEERKGMDDTLKL 1237
 DB 1190 EEEIISDEDEI 1202
 RESULT 2
 PDS5_YEAST
 ID_PDS5_YEAST STANDARD; PRT; 1277 AA.
 AC Q04264; Q04780;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Sister chromatid cohesion protein Pds5 (Precocious dissociation of
 DE sisters protein 5).
 GN PDS5 OR YMR076C OR YM9582.01C OR YM9916.15C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RX PubMed-9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagsels K., Iye G., Moule S., Odell C., Pearson D., Rajadream M.A.,

RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT XIII.";
 RL Nature 387:90-93(1997).
 RN
 RP FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20517447; PubMed=11062262;
 RA Hartman T., Stead K., Koshland D., Guacci V.;
 RT "Pds5p is an essential chromosomal protein required for both sister
 RT chromatid cohesion and condensation in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 151:613-626(2000).
 [3]
 RN
 RP FUNCTION.
 RX MEDLINE=21065936; PubMed=11137006;
 RA Panlaza S., Tanaka T., Hochwagen A., Eisenhaber F., Nasmyth K.;
 RT "Pds5 cooperates with cohesin in maintaining sister chromatid
 RT cohesion.";
 RL Curr. Biol. 10:1557-1564(2000).
 CC -1- FUNCTION: Essential for the establishment and maintenance of
 CC sister chromatid cohesion at centromere proximal and distal
 CC regions during S phase. Also required for chromosomal
 CC condensation.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 249259; CAAB9222.1; -
 DR EMBL; 248952; CAAB8801.1; -
 DR PIR; S5451; S5451.
 DR SGD; S0004681; PDS5.
 DR GO; GO:0005708; C:mitotic chromosome, IDA.
 DR GO; GO:0007076; P:mitotic chromosome condensation; IMP.
 DR GO; GO:0007064; P:mitotic sister chromatid cohesion; IMP.
 KW Mitosis; Nuclear protein.
 SO SEQUENCE 1277 AA; 147040 MW; 9DF40A5274FD9623 CRC64;

Query Match 7.7%; Score 557; DB 1; Length 1277;
 Best Local Similarity 20.3%; Pred. No. 2.7e-18;
 Matches 272; Conservative 267; Mismatches 611; Indels 190; Gaps 46;

12 KITYPGVKEISDK -ISKEMVRRLKMYKTFMDMDQSEEEKL--YLNALHLASDF 68
 8 KLEKNSPIITSDQLISTNELDLRLKALHEELASLDQDNTDLTGDKYDA--LVSRL 64
 69 LKHGKDVRLVACCIADIFRIYAPAPYSPDKLKDIMEFTRQKLEDTKSPQFNRY 128
 65 LKHKVGIRATFACCLSDLRKLAPDAPTTDA-QLTDIRKLYVSQPEQIGDDENGYHIO 123
 129 FYLEENIAMVKSYNICFELEDSNEIFTQLYR-----TFSVINGNHQKVMHM 177
 124 TYLIRKLEYSRIVLLADPLSSNNLLIEFHIFYPDNKSFPARLENVIGS----- 173
 178 VDLMSIICEGTVSQOELLDTVL-----VNLVPAHKNLNQ-AVDLAKALLKRTAQAI 229
 174 --TIGEVISEFDSVPLEVLRLLFNKFLTYNPNEIPEGNAVTSDCGEVSLIICDYSNNM 231
 230 EBYITTFNOVLML-----GKTSISDLSEHVFDLIELYNIDSHLLSVLPQLEFKSN 284
 232 SNHLTKYSEIHEATNDNNSRLTLVYVVKLKLVLRLMEYPELNNAYIGFIYHELSE 291
 285 DNEERLQVYKLLAKMFGA-KSELASQNKPLMQCYLGRPNIDHVPRLCEVAFASHCLMN 343
 292 NELFKEATKTLIGQILTSYDLNFVSTHSDTFKAMISLADISPDVAVEMWESIQIAT 351
 344 HPDLAKDLTEYLAKVSHPEEAIIRDVYST--VTAAKQIL--LVNDHLNLYREPT 397
 352 REDISKEINQALAKTFIDSPRVRTSVIFNKVPYTEIMKNITKAITSLLHARE-- 409

QY 398 LDKRWVRKEAMGLAQIYKKAQO---SAAGKDAKQJAMIKDKLLHYONSIDRLL 454
 410 --KHKVRELCINTAKKFTYSNSLNEIERKYONKEIWEIIDTIPSTLYNXYINDLNINQ 467
 QY 455 VERIEAQYVPHNLLETTERMKCLYLYATPLDANAVKALNEMKQNLRLHQYKDLLIR 514
 468 VQSVIEFYLPEEPDNDKHVHRLVLSHFDDKKAFTSPFAFNARQIKISFALSKEYIDFSK 527
 QY 515 QKRTASVAKAISKVVV-----ITRNLPDPCAOPPMKFTYVLEDEDEKIRKQEV 565
 528 FLNDESMSSSOGPIYMNKYNOTLQWLASGLSDSTKAIDALETIQF--NDERIFYLNA 585
 QY 566 LVSPTSCQKQAEQVREITKRLGNPK-----QPTNPFLMIKFLLEIRAPVH 612
 586 CYTNDIPELTFKNCYELVSKIQTPOLFKRYNISTGASIMPRD-IAKVQIILLFRASPII 644
 Db 645 YVNSNISVNLNSNNS-----DAKQDLKRRILDISKYNPLFKDQIRT 689
 QY 670 FESLACL-----KMDDEKVAEALQIFKNTGS-KIEDFPIIRALLPVLHNSKGP 722
 690 LKTIILKDLDDPAEKNDNLSEALKTLYLKASKTLYKQVDFPD--TFEFTKLYDAVESK 747
 QY 723 PROAKYAIICIAHAFSSKETOFAOI--PEPLKSLDPENLEHLITPLTIGHIALPAD 779
 748 PETTKATKTLI-ALSKAEETLKIKIRLPDLQDKQFTSHIYIMLFFKFPVLAN 806
 QY 780 QFAAPKSVAVTFYVKDLMLNDRLPCKRTTKL-WPDEEVSPETVW---KIQAIKMVR 834
 807 D-----STDIISYLKEVLLSNQVGDSSKEIDWDSLSLPTKYSAGIKNFYTLKLF 861
 QY 835 WLGM-----KNHNSSGT-SFLRLTLTLLHSDGL-TPQGIKSRDMS---RLPLAG 883
 862 KLRSTAPDVRDELSEFTKTKLFEYTLASGGLISEFNKEFYTPSNYOTKLCVAG 921
 QY 884 SAIVKLAOPCYHEIITTELOYOLCALAINDECYQROYAPQKILHGKLSRLRLPEYMAIC 943
 922 IQVLKARISNNLNFIPKSDIIRKLNLYVDESILPRKRTPLBQDKQYVANELSIRFLPV 981
 QY 944 ALCANDPYKERRAARQCLVKNINVRREYLKQHAANVSEKLSLLEPVVYPTIHLANDP 1003
 982 FETAAVEPDELTTTKIWI--NFTGLKSFKR-----GTIFERAPRLIHAIHNP 1030
 Db 1004 DVYKVDIF-----QLADVKECLMFVLEIIMAKNENSHAIRKNEVIK-----QTK 1051
 1031 DIVGGIDSEGDAYLNAITTAIDYLLFEFDSIAAOENFSLLYLSERKNYQDKLVEDET 1090
 QY 1052 DAQGPDDAK-----MNEKLYTCDVANIIIMS-----KSTYS-----LESPPKPVLA 1095
 1091 EEEGQKEAPKPKKHRYGQKMTIIGELSQMILNLNKEKKNMQHSAYPGKLNLPDLFKP- 1149
 QY 1096 RFTQPDKNFSMTKNYLIPPEMKSFPTPKRPTTNVLGAVNKPLSSAGKOSQTKSSRMETV 1155
 1150 -FATVQEAQLS-FKTYIPESLT-----EKIGNNKAKIGRLHNSQTRQRLQKRLLAH 1201
 QY 1156 SNASSSSNPSSPGRIKGRIDSSEMDHSEMEDYTMSPLPGKSD-KRDDSILVRSELEK 1214
 1202 ENNESQKKKKVHNAHQADDEGDRESDSDSYSPSNKNETKKGHEINIV---MKKL 1258
 QY 1215 RGRKTPVYEOEEKLGMDL 1234
 Db 1259 RVRKEVDYKDDSD---DDI 1274

RESULT 3
 BPEA_HUMAN
 ID BPEA_HUMAN STANDARD: PRT: 5171 AA.
 AC 094833; 08N178; 08N8J3; 08WKX9; 096AK9; 096Q05; 09H555;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 28-FEB-2003 (Rel. 41; Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last annotation update)
 DE Bullous pemphigoid antigen 1, isoforms 6/9/10 (Trabeculin-beta)

DE (Bullous pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein)
 DE (Dystonia musculorum protein).
 DE BPAG1 OR DMH OR DT OR KIA00728.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 6), AND TISSUE SPECIFICITY.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=21839111; PubMed=11751855;
 RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;
 RT "Novel alternative splicing of BPAG1 (bullous pemphigoid antigen 1)
 RT including the domain structure closely related to MACF (microtubule
 RT actin cross-linking factor)."
 RL J. Biol. Chem. 277:6682-6687(2002).
 RN [2]
 RP SEQUENCE OF 1342-5171 FROM N.A. (ISOFORM 10).
 RC TISSUE=Dundenum;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vyllaion D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fehey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 1702-4156 FROM N.A. (ISOFORM 6).
 RC TISSUE=Brain, Placenta, and Tongue;
 RA Niimiya K., Nagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
 RA Kusuta N., Sato K., Tanikawa M., Yamazaki M., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Isogai T., Nishi T., Ota T.,
 RA Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Kikuchi H., Masuno Y.,
 RA Nagai K.;
 RT "NEDD human cDNA sequencing project."
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 3535-5171 FROM N.A. (ISOFORM 6).
 RC TISSUE=Brain;
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones."
 RL DNA Res. 9:99-106(2002).
 RN [5]
 RP SEQUENCE OF 4107-5171 FROM N.A. (ISOFORM 6).
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286(1998).
 RN [6]
 RP SEQUENCE OF 4031-5171 FROM N.A. (ISOFORM 9).

RA Smith M.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP DISEASE.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=92011493; PubMed=1717441;
 RA Sawamura D., Li K., Chu M.-L., Uitto J.;
 RT "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
 RT deduced from cloned cDNAs predict biologically important peptide
 RT segments and protein domains."
 RL J. Biol. Chem. 266:17784-17790(1991).
 CC -1- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
 CC intermediate filaments to the inner plaque of hemidesmosomes. The
 CC proteins may self-aggregate to form filaments or a two-dimensional
 CC mesh (By similarity).
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named Isoforms=10;
 CC Name=6; Synonyms=EA;
 CC IsoId=Q94833-2; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q03001-1; Sequence=External;
 CC Name=2;
 CC IsoId=Q03001-2; Sequence=External;
 CC Name=3; Synonyms=1e;
 CC IsoId=Q03001-3; Sequence=External;
 CC Name=4;
 CC IsoId=Q03001-4; Sequence=External;
 CC Name=5;
 CC IsoId=Q03001-5; Sequence=External;
 CC Name=7; Synonyms=EB;
 CC IsoId=Q8WXK8-2; Sequence=External;
 CC Name=8;
 CC IsoId=Q03001-6; Sequence=External;
 CC Name=9;
 CC IsoId=Q94833-3; Sequence=VSP_005068, VSP_005069;
 CC Name=10;
 CC IsoId=Q94833-1; Sequence=VSP_005066, VSP_005067;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and
 CC cultured keratinocytes.
 CC -1- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
 CC [MIM:600088], an autoimmune subepithelial skin blistering disease.
 CC -1- SIMILARITY: Belongs to the plakins or cytolinker family.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- SIMILARITY: Contains 29 spectrin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF400226; AAL62061.1; -;
 DR EMBL: BC016991; AAH16991.1; -;
 DR EMBL: AK05189; BAB70870.1; -;
 DR EMBL: AK094883; BAC04449.1; ALT_INIT.
 DR EMBL: AK096713; BAC04448.1; ALT_INIT.
 DR EMBL: AB018271; BAA34448.2; -;
 DR EMBL: AL137008; CAC12899.1; -;
 DR HSSP: P02631; IRRO.
 DR Genev: HGNC:1090; BPAG1.
 DR MIM: 113810; -;
 DR MIM: 600088; -;
 DR GO: GO:0005200; F:structural constituent of cytoskeleton; ISS.
 DR GO: GO:0045104; P:intermediate filament cytoskeleton organiza. ; ISS.
 DR Interpro: IPR002048; EF-hand.
 DR Interpro: IPR003108; GAS2.
 DR Interpro: IPR002017; Spectrin.

DR Pfam: PF00036; efnand; 2.
DR Pfam: PF02187; GAS2; 1.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00435; spectrin; 28.
DR ProDom: PD000012; EF-hand; 1.
DR SMART; SM00054; Eph; 2.
DR SMART; SM00243; GAS2; 1.
DR PROSITE; PS00018; EF_HAND; 2.
KW Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW Calcium-binding; Alternative splicing.
FT REPEAT 264 341 SPECTRIN 1.
FT REPEAT 349 444 SPECTRIN 2.
FT REPEAT 450 550 SPECTRIN 3.
FT DOMAIN 564 616 SH3.
FT REPEAT 935 1031 SPECTRIN 4.
FT REPEAT 1481 1581 SPECTRIN 5.
FT REPEAT 1715 1814 SPECTRIN 6.
FT REPEAT 1930 2008 SPECTRIN 7.
FT REPEAT 2071 2174 SPECTRIN 8.
FT REPEAT 2184 2282 SPECTRIN 9.
FT REPEAT 2294 2382 SPECTRIN 10.
FT REPEAT 2404 2502 SPECTRIN 11.
FT REPEAT 2513 2611 SPECTRIN 12.
FT REPEAT 2839 2940 SPECTRIN 13.
FT REPEAT 2950 3047 SPECTRIN 14.
FT REPEAT 3059 3156 SPECTRIN 15.
FT REPEAT 3168 3258 SPECTRIN 16.
FT REPEAT 3277 3376 SPECTRIN 17.
FT REPEAT 3386 3484 SPECTRIN 18.
FT REPEAT 3495 3594 SPECTRIN 19.
FT REPEAT 3716 3816 SPECTRIN 20.
FT REPEAT 3925 3925 SPECTRIN 21.
FT REPEAT 3935 4035 SPECTRIN 22.
FT REPEAT 4044 4142 SPECTRIN 23.
FT REPEAT 4153 4250 SPECTRIN 24.
FT REPEAT 4262 4362 SPECTRIN 25.
FT REPEAT 4372 4470 SPECTRIN 26.
FT REPEAT 4480 4578 SPECTRIN 27.
FT REPEAT 4596 4684 SPECTRIN 28.
FT REPEAT 4732 4790 SPECTRIN 29.
FT CA_BIND 4798 4810 EF_HAND 1 (POTENTIAL).
FT CA_BIND 4834 4846 EF_HAND 2 (POTENTIAL).
FT DOMAIN 126 160 COILED COIL (POTENTIAL).
FT DOMAIN 312 377 COILED COIL (POTENTIAL).
FT DOMAIN 402 447 COILED COIL (POTENTIAL).
FT DOMAIN 511 546 COILED COIL (POTENTIAL).
FT DOMAIN 706 813 COILED COIL (POTENTIAL).
Query Match 3.3%; Score 239.5; DB 1; Length 5171;
Best Local Similarity 18.5%; Pred. No. 0.006; Mismatches 448; Indels 467; Gaps 66;
Matches 263; Conservative 241; Mismatches 448; Indels 467; Gaps 66;
QY 22 ISDKSKSEWVRRLKMYVTFMDMDQSEEEKELYLNLAHLSDFELKHPGKDYRLAVA 81
DB 2395 VKEEVLKAEKLSRQGLSKSEFDEOKAE-----NHVQHLSQACASHQ----- 2438
QY 82 CCLADIENIYAEAPYTSBPKLIDFIMFTTRQLKGLDETKSPQENRYFYLLENIAMVKS 141
DB 2439 -----FOOMSROFQAMLDTRKKEQN-----KSH 2461
QY 142 NICEPELEDSNEIFQLYRTLPFSVINNGH-----QKVMHNVADMSSITICGDRV---- 191
DB 2462 PISAKLD-----VLESILKDHKDFSKTLTAOSHMY-----EKTIAEGENILKLT 2505
QY 192 -----SOELDIYLVNLYVPAHKNL-----NKQAYDLAKAL-LKRTAQALBEYITTFNQ 239
DB 2506 QGSEKALQLOLQNTIKTWMDFPNKQVKEKREKLSLEKALKYKKEOVETLMPWIDKCOQN 2565
QY 240 VIALM-----GKTSIS-----DLSEHVFDLILELYNIDSLLSVLPOLERKLSN 284
DB 2566 LEEIKFCLDPAEGENSIKLSKLOKEMDH-FGMV-ELINNTANSILSYC-EDRGVAVTD 2622

QY 285 DNEERLOVYKLLAKMGAND-----SELASQNKPLMOCYLCGRNDIHVP-- 329
DB 2623 ENKSLILQKVDVTEQHSKSKFCLENNMTQKFEQHEVSKESKROLQCARQOL-DIHSLGS 2681
QY 330 -----RLECVFASHCLMHNHPLADLTLEYLKVSRHDEBEAIRHYVISTYAAK 380
DB 2682 QAVSNKYTLMTLOOQSLQALKHQVYDLAKRLQDLAVEASDSKGT--SDVLLQVETIAOE 2739
QY 381 DLLVNDHLNVRRETLTKRMVRVREAMVGLAQIKYKVALSAAGKAADKAOLAMTKDL 440
DB 2740 -----HSTLSQVDEKCSFLEKL 2758
QY 441 LHI-YYONSIDRLIVERIFEAQVWPHNLETTERMKCLYLYATLD--LNAKVALNEMK 497
DB 2759 QGIGHFRNTI-----REMFQ-----FAEPDELDMSAPVQ---- 2789
QY 498 CONLHRHOVKLLDLIKQPKTASVKAIPSKVAVITRNLPDGKAODEPKKFTQVLEDE 557
DB 2790 -----RDAGETLOKQKET--IKAFLEKLEALMASNDANKTKCMMLATEETSPDLV 2837
QY 558 KIRKOLEVLYSPQSC-----KQAGCVREITKKLGKNKOPTNPFLEMIFLELRI 608
DB 2838 GIKRDEAL-SKQCNKLLDRAQAREQVGTIKRLEEFYSKLE-----PSILQRA 2888
QY 609 -----AFVHIDTESIALI-----KQVKSIDGTADDED--EGV-----PTDQ 644
DB 2889 EHEHESQGVGMEETETINQLMKFKVQKEBELIEPLQKQODVNWGLQGLQSAKSTSQ 2948
QY 645 AIRAGLEILKV--LSFTNPISFHSATFESSLACLKMDKQVAAALQIFKNTGSKIED 702
DB 2949 GLEHLDIDVNAEMKTLNKKVAAQRAAOLQALHLHGCFQD--ALESLISMVAVTEELVANO 3006
QY 703 PPHISALLPVLIHNSKKRPPROAKYAINCHAIITSKET-----QFOQIEPLMK--- 753
DB 3007 KP-----PSAEFKVKAQIQEQKILQRLDLDKRSSTVEVIRREGKIAFTAPADKVKI 3059
QY 754 -----SLDPSNLHLITPLVT-----IGHILAPDOF--AAPKSVATFYKDLMDNR 802
DB 3060 LKQSLIDSRWELALINKATRRKOLEGISVA-QOFHELEPLNEMLTT--IEKRLVNC 3116
QY 803 LPKQTKTKLMPDDEEVSPEYVWKIQAIKMMVWMLQMKNNHSK-----SGTSTRLTLTT 856
DB 3117 PIGTQASKL--EEOQAQKALEDDII-----NNKHHLQAVSIGSLKVLSS 3161
QY 857 ILSHSGDLTEQCKISKPDMSR-----LRLAAGSAYVKAQOEPCHYEITILEQYOL--AL 909
DB 3162 --REKDMVQ-----SKLDFSQVYVYIEIQKSHSRSELLQQAQCNMKIFEGEDEVELMNVLN 3215
QY 910 AINDCC--YQVRQVFAQKHLKGLSLRLPLEYMAICALCAKDPVERBRAHAROCYKNN 967
DB 3216 EYHDKLSKLSVODYSTEGMLKQOSELRV-----LQEDILRKQNVQDALLNLGL- 3263
QY 968 VREERYLKQHA-----VSEKLSLPEYVVPYTHLHADHPYVAVODIEQL-DVKECL 1021
DB 3264 ---ELLKQTTGDEVILIQCKLEAIRKY-----KDTIKLSTQVAKTL 3302
QY 1022 WFLVELIQAANNENSHAFIRKRVENIKQTKDAQGDDAKMNEKLYTVCDVANNIIMSST 1081
DB 3303 EQALO--LARRLHSTH-----BELCTWIDKVEVELISYET 3335
QY 1082 TVSLSPKQPVLPARFPTQPDKNFSNTKNVILPPEMKSFPTPKPPTVNLGAVNKPPLSA 1141
DB 3336 -----QVLKGEASQAQMR-----PKELK----- 3354
QY 1142 GKQSQTKSRMETVSNASSSSNPSPGRIGRLKGLIDSSSEMDSHSENYDTMSSPLPGKSDKR 1201
DB 3355 -KEAKNNKALLDLSLNEVSALLLELVPMWRARELE--KMAVADENERYRIVSDTITQKVEEV 3411
QY 1202 DSDSLVRSLEKPRGRKTPYVTEQBEKL-GMDDLTKIVQEQPKRSQ-----NSR 1250
DB 3412 DAAILRSQFPQADAELSMITETBKRLMSLADI-RLRQDQVSAOLQOVQKTFYMEILRRH 3470
QY 1251 -----KRGH--TASESDQOQWPEKRLKEDILNED 1279


```

Db 1343 -----SKLSQSLFEQVNSLRDA-----KNVLNLSSEANNAIELOAKVQ 1385
OY 1127 TTNVCAVANKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIGRLDSSMDHSENED 1186
Db 1386 GNNQLEAIRKTLQEDAEKASRELQAKLE-----ESTSYESTINCL-----NBE 1428
OY 1187 YTMSSPLPGKSKSDKDDSLVSELEKPRG-RKKTPVYQEEKLCMDLTIKLYQOKPRG 1245
Db 1429 IT-----TLKELEKQRIQOOLQATVSAHQ-----NDLSNIVESSK--- 1465
OY 1246 SQRSSRKRHTASESDQWPEEK--RLKEDILNEDQNSPP-----KKGKRGRRPKP 1296
Db 1466 -----KSFEEKIKFKIKETQEVNKKILAEQERLNQPSNIMETKKKWESEHDE 1516
OY 1297 LGGGPKKEEPTMK-----TSKGSK---KSGPPAPEEEEEEER-----OSG----- 1335
Db 1517 VSQKIREAEALKKRIRLPTEKIKIKIERKKEELEKEEKEVEERIKSMQSGEIDVYL 1576
OY 1336 -----NTEOKSKSKOHVRSRAQOARSS-PSSASLESTQ 1368
Db 1577 RKOLEAKVOKEKOLENEYNNKLOBELKDVPHSHSIDDE 1616

RESULT 5
CENE_HUMAN STANDARD: PRT: 2663 AA.
ID CENE_HUMAN
AC 002224:
DC 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RX Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RX Thirmer D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RX Chan G.K.F., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
RP KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
RP OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
RP AND/OR SPINDLE ELONGATION.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: 215005; CAA78727.1; -
DR PIR: S28261; S28261.
DR HSSP: P17119; 3KAR.
DR Genew: HGNC:1856; CENPE.
DR CK: Q02224; -
DR MIM: 117143; -
DR GO: GO:0005699; C:kinetochore; TAS.
DR GO: GO:0005634; C:nucleus; TAS.
DR GO: GO:0008350; F:kinetochore motor activity; TAS.
DR GO: GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO: GO:0007079; P:mitotic chromosome movement; TAS.
DR GO: GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin_1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA: 312087 MW: CEC138B0C8C8CB8 CRC64;

Query Match 3.2%; Score 229; DB 1; Length 2663;
Best Local Similarity 18.2%; Pred. No. 0.0077;
Matches 275; Conservative 268; Mismatches 537; Indels 434; Gaps 64;

OY 17 PGVKEISDKISKENEVRRLKIVKVTQMDDQDSE-----EKELYLNLAHLASD 66
Db 333 PYVNEVS--TDEALLKRYR--KEIMDLKOLEVSLQETRAQAMEKD--OLADLLEK 383
OY 67 FFLKHPKGDVRLVAVACCLADIFRIYAPAPTSPPKLDIF-MFTIRDLKGL-EDTKSPQ 124
Db 384 DLQK-----VQNKETINLTMLTSSSTLTQOELKRR 417
OY 125 ENRYFYLLFNIAVWYSNICEFELDSNEITFOLYTLTSVINGNHQVHMHVMDLMSI 184
Db 418 KRYVWCICLKINKKMSN-----YADQNIPIPN-ITTYTHKLSINLKEI 461
OY 185 ---ICEGPTVSOELDVLV-VNLVPAHKNLKQAVDLAKALKIKRTAQAIPEYITTFNOV 240
Db 462 DESVCSSEDSVFSNTLDTLSEIEMNPATKILNDENI----- 497
OY 241 LMLGKTSISDSEHFIDLILELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLAKMF 300
Db 498 -----SEINSLRADYDNLVL--DYQLRTREKEMELKEKNDLDFEALERTK-- 545
OY 301 GAKDSELASQMKPLMQCYLGRFNIDIHVIRLECVAFASHCLMNHPLDAKDLTEYLKVRSH 360
Db 546 --KQDEK-----LIHEISNLKNLYKIHREYVNO 571
OY 361 DPEAIRHDVIVSYTAAKKDILLVNDHLNLFVRERTLDKRRRKEMAMGLAQIYKYYA 420
Db 572 DLENEL-----SSKVELLREKEQIKKLEQYIDSK--LENIKMPLS-----YS 613
OY 421 LOSAAGKAAQAIAMIKKILHIYQNSIDRLLVERIFAQYVPHNLETTERMCLYLL 480
Db 614 LESTIEDPKOMQOTLPDAFTV-----ALDAKRESAPLSENLELKEKMEKELATY 661
OY 481 YATLDLNAVKALNEMMKCONLRLHQVKDLDLIRKPKDASVKAIFSKVAVYTR----- 534
Db 662 YKQNE-----NDIQLYQSLLEAKKKMQVDLEK-----LQSAFNRITLTSIDSKV 708
OY 535 -----NLPDGKADPFKKTFTYVLEDEKTRKOLEVAVSPSCQAQAGCVREITKLG 588
Db 709 PKDLICNLEGGKITDLOKELNKEVEENALREEV-ILLS--ELKSLSPSEVERLRKEI- 763

```



```

Db 243 NNIIPLTNTKNNVEYDOKENILIVNIESIIEICIPHLQIAODPTLLSSSEKKNPVYI----- 298
QY 85 ADIFITVAPAPYTS-----PD---KLDIF-----MFTTR 112
Db 299 ---RLVYOIVRFESFSAINSFKIVKLTGRPDVVKLVYIRMTTGALRNSNKIITTA 354
QY 113 QLKGEJEDRSQOFNFYFLENI-AMVKSYNICEFLEDSNEIFTOYLTPLFSVINGNHQ 171
Db 355 QVSEFIRDEK---FGTFFLSNEETKPIISTFTEIMEINSHNLIYEKL-----LIRGFLS 405
QY 172 RVHMHVMDJMS-----ICBGDTVSOELDTVLVNLVPAKNNKQAYDAKAL 222
Db 406 KYPKRIETVTSWTLGEVLPRIIIGDEIYSKILITSIVLL-----ELL 450
QY 223 KRTAOAIPEYITTFPNOVLAMGKTSISDSEHVPFLIELY---NID---SHLLSVIP 275
Db 451 KKCIDFVDEHERIY--QCIMLSPVCETIPEKFLSKLPINSYDSANLMDKVTIGHLLTQOIK 508
QY 276 OLEFKLGSNDNEERLOV-VKLAKMF--GAKDSELASONKPLM-----QCYLGRPNDIHV 327
Db 509 ---NYIVAKNDKIAMDLMSMTGLIYDSGKRYDITLSESNKWMPDLNLCFINN---HP 562
QY 328 PIRLECVR--FASHCL-----MNHPLAKDLTEYLK-----VRSHEEALRHVY 370
Db 563 KTRLMISIKVMRIITFYCTIKISQKQEGNKSLSLRTPQWTLRPVNDPSAREGIIYHL 622
QY 371 IVSIYTA--AKKDIL-----LVNDHLNLFVRERTIDKRRVR-----KEA 408
Db 623 LGVYTAFTSNKNLSTDMFELEMDHLITPIYEDYFKDISHLQVLEFTVLHLIGKVA 682
QY 409 MMGLAQIYKKY-----ALQSAAGKQAK--QJAWIK--DKLHIYYONSIDRLVY 455
Db 683 DYALEKRTKKHHPMSVIASBEGVAKLDISSLPQI--IKREYDKIMKVFQI-----V 733
QY 456 ERIFAQVWPVPHNLETTERMKCLYLLATFLDLNAVALNEMKQCNLLRHQVKKOLLDLIKO 515
Db 734 EVAISNVMLADHLILTS-----LKHLPED----- 757
QY 516 PRTDASVKAIFSKVAVITRNLPDPGKADPFMKFTOVL-----EDEKIRKOLE 564
Db 758 RKOQTHLESFSSLLIKVQNNKDPPIFRDFEGAVSSFEVYFLDLFLKRNDSLVNPNQO 817
QY 565 VLVSPFCSQKQAGCVRREITTKIGNPKOPTNPFLEMIKFL-----ERIA 609
Db 818 ISKVGISQGNMTLDLKVYIRKARN---ETSEFLIEKFLDLDDKTEVYAQNWGSTLL 874
QY 610 PVHIDTESIALIKOVNKSIDGTADDEDECVPTQAIRAGLELLKVLSTFPHISHAET 669
Db 875 PNISFREFOSIANIVNK-----VENENSIENFDLCLKLSF--PVNL----- 915
QY 670 FESLLACLKMDKEKVAEALQIFKNTGSKIEEDF-----PHIRALLPVLHHS 718
Db 916 FTLLHVSMNSNNNFYFQSYVSKNE-NKLVNDLITLKTSLPGNPELSGGLPVL----- 970
QY 719 KKGPRQAKY--AIHCIIA-----IFSSKE- 741
Db 971 ---RRMKFMDILEYCIHSNPNLNSIPDLNSDLLKLPSSRSASYFANIKIKCSQO 1025
QY 742 -----TOFAQIFPEPLKSLDPSNLE-HLITPLVTIGHIALLPADOFA 782
Db 1026 ITLVWMLLKGOOLEONFONFEISEIVNLVNASD-SELEKSEIIRLEL--HLAMNPTE-- 1079
QY 783 APWKSVAWTFIYKDLMLN--DRLPGRKTKTKLVPRBEVSPREMNVIQAIKMMVNRMLGK 840
Db 1080 -PLEFGLNFCIKNNMAHDLDEFEGNMTSEVLP--KISPELLIKLITYK----- 1125
QY 841 NNHSKSGTSLRLTLTIHSDGD---LTEQKISKPDMSRLIAGAIVYKLAQEPCH 896
Db 1126 ---ERKNGKILAVIEKIENGDDDIYELLEKIITIOKE-----IOLKLEKP--- 1170
QY 897 EITTEQYOLCALAINDECYO-----VROVFAQKLHKGLSRLPLEYMAICALCAKD- 950
Db 1171 ---LLVEFLNFSVSNMQKHKKSTWMLRELVLVLTPLPSRSAAKKFMSLISILPPNEN 1226

```

```

QY 951 -----VERRAHARQCLVKNINVRRELYKO-----HAASEKLSTL 986
Db 1227 YQIDMVNLLIDLKSHNKKFKDRYY--NATLKTIG---KWIQSGVHOGDSSEKIEA 1281
QY 987 LPE---YV-----VPYTIHLADHPDVYKODIEOLKDYEC 1020
Db 1282 IPDFTSMYIIPCSESNKLSNLOARKVDSODIQVPAT-QGKKEPSSIQSSQISAND-SDS 1339
QY 1021 LMFVLEIIMAKNENSHA-----FIRKNENIKQTK-----D 1052
Db 1340 ISLKNATAMNSSQOESHANRSKSIDETLEEVNDSIEHIDQOMKSTOLDKKNVANHSNIC 1399
QY 1053 AOGPDAMNNEKLYVVCVAMNIMSKSTYSLESKPOVLPARFPOTPKRFSNTKNYL 1112
Db 1400 STKSDEVDYTE-LHESIDIOSSEVNAVOPIEVLTSELKAVTNRSKITNDPHNVAVSONPL 1458
QY 1113 PPEMKSFTTPGKPKT-----TNVLGAVNKPPLSSA----- 1141
Db 1459 KRPSKETPTSENRKSGHETWVVLVSEBQAVSPSSDYICTNIKSIANDESSLARNSIK 1518
QY 1142 -----GQSOTKSSRMETVSNASSSSNPSSGRITKGRLDSE----- 1178
Db 1519 VETNCNENSLVNTLDDOQITIKEDKGQVEHVORQOENQSNKJNSKSTODNIAOYKS 1578
QY 1179 ---MDHSENEYTMSPLPGKSKDKRDSDLRSE-----LEKPRGRKTPYTEQBE 1227
Db 1579 VKKARPNNGENNDIACNV-----EQASPVRNEYPGGIDIPSTIILLNSKOTE 1628
QY 1228 KLGMDLTK-----LVQEQPKGORSRRKRGHTASEDEQOEMPERKLEJLLEN-- 1277
Db 1629 KSKVDLSDSEDEHGTAVQEKHQGAINSRNKNNDRMOSTPIQGTSEESREVVMEEGIN 1688
QY 1278 ---EDQNSPPKKGRGRRPKPLGG-----PKREPTAKTSKKGSKSG 1320
Db 1689 VLEDSGTGCELKNKLG---PLKGDKANINDEVPVEENVRDGFGLSMHAYSKETG 1744
QY 1321 PPAPEEPEEE-----ROSGNEOKSKSKOHVSRRAQOARSPRESS 1362
Db 1745 ---LEQPEVADISVLPRIPIPLFNSLKMGSQIKKELKRLQRLNELMPPSPPRM 1799
QY 1363 ALESTQSTPQKGRGRPSKT 1381
Db 1800 T-ENTNINAMNGLDTPVPT 1817

RESULT 7
ALM1_SCHPO
ID ALM1_SCHPO STANDARD: PRT: 1727 AA.
AC 09UTK5; 01313; 09UTR8;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayes S., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Buckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

```


QY 1145 SOTKSSRMETVSNASSSSNPSPGRIGRLDSSSEMDHSENDYTMSSPLPGKSPD----- 1199
 Db 1182 RQLLEFKKRLKRVESNN-----RVIAIDLRGRTTEKDVSTSSVSVERNYLMNV 1232
 QY 1200 --KRDDSDLVNSELKPRGRKRTPTVEQEEK-----LGMDDLTKLYOEQKP- 1243
 Db 1233 ALNENKNSLHLENIE-----RNEEVITELREKIEPLTKDLANFRLNKKEQLSOLQTEKAAV 1288
 QY 1244 ---KGSQSRKR-----GHTASESDQCPPEKRLKEDLLENEDQNSPPKGGKR 1290
 Db 1289 KKLNSNNEEYKRNHOELLNLNSSTSTSDASRLKNELVSKENILEELNDEI----- 1340
 QY 1291 GRPEKPLGGCTPKPEPTKTKSGSKKSGCPAPEDEEEROSGNTPE----- 1338
 Db 1341 -----GLHKEE--LETKKSKSEDLNENRAONOSKIEQLEKNTKLAAMRTKYEQV 1389
 QY 1339 -QKSKSOHVRVSRRAQOAPESPESASIESTGTPQKGRGPKTP---SPSOP 1387
 Db 1390 VNKSLEKHQIRQOLSKRTSELEKVAECQHLNEQ--LNKPSATPTATTOSEP 1440

RESULT 8

NFH_HUMAN STANDARD; PRT; 1020 AA.
 AC P12036;
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 GN NEFH OR NFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86328981; PubMed=3138108;
 RA Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 subunit (NF-H) and the gene encoding it.";
 RL EMBO J. 7:1947-1955(1988).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC OBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X15306; CAA33366.1; -;
 DR EMBL: X15307; CAA33366.1; JOINED.
 DR EMBL: X15308; CAA33366.1; JOINED.
 DR EMBL: X15309; CAA33366.1; JOINED.
 DR PIR: S00979; QFHHU.
 DR Genew: HGNC:7737; NEFH.
 DR MIM: 162230; -;
 DR GO: GO:0005883; C:neurofilament; NAS.

DR GO: GO:0007399; P:neurogenesis; NAS.
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 FT DOMAIN 1 100 HEAD.
 FT DOMAIN 101 413 ROD.
 FT DOMAIN 414 1020 TAIL.
 FT DOMAIN 101 132 COIL 1A.
 FT DOMAIN 133 145 LINKER 1.
 FT DOMAIN 146 244 COIL 1B.
 FT DOMAIN 245 266 LINKER 12.
 FT DOMAIN 267 288 COIL 2A.
 FT DOMAIN 289 292 LINKER 2.
 FT DOMAIN 293 413 COIL 2B.
 SO SEQUENCE 1020 AA; 11779C9CB3DCFD14 CRC64;

Query Match

3.0%; Score 217.5; DB 1; Length 1020;

Best Local Similarity 19.2%; Pred. No. 0.0074;
 Matches 206; Conservative 162; Mismatches 426; Indels 281; Gaps 44;

QY 377 AAKRDILLVNDHLN---FVREPTLDKRWVRKRAMGLAQIYKRYALQSAGK-DAKQ 432
 Db 162 AARGQLRLEQSHLELDIAHVQR-LDDEARQREBA-EAARALARFAQEAARAVDQKK 219
 QY 433 IAWTKDK--LHIVYQNSIDRLIVERIFAQYVPHNLEPTERMKC-----LYLVATID 485
 Db 220 AQAQLEEGCYRRHNOEEVGLAQIQSGAAQAOQAQOAEYTDALCKDVTSLREIRQLE 279
 QY 486 LNAVKAL--NEMMKCONLIRHQVLDLIDLIQPTDA-----SVKAFSKVWVITRNL 537
 Db 280 GHAVOSTIQSEFWVRVL-----DRLESAKQVMTDARSQAQOEITETVRQQLARTTELE 333
 QY 538 DPGKAQDPMKKFTQVLEDEKIRKQLEVLSPTSCQKAGCVCVREITKLCNPKQPTPF 597
 Db 334 ALKSTKDSLERSQSELED---RHQADI--ASYQRA--IQQLDAELNRTWEMAAQ 381
 QY 598 LEMIKELLERLAPAHIDFESIALIKQVNSIDGTADE-----DEGVPTDQA 645
 Db 382 LREYQDLN--VKMALDIE-----IAAYRKLDLG--EKGIGGPIPFSLPEGLPKIPS 431
 QY 646 IRAGL-----ELKLVSTFHPISFHSATFESLCAKMDKVAEALQITKN----- 694
 Db 432 VSTHIVKSEKIKVKESEKETYIVEQETQTEVTEVEDEEKAKEEGKEEGEE 491
 QY 695 ---TSGKTEEDPPIHRSALPYLVHKKSGP-----PRAKYAIHCIAIFPSKETQF 744
 Db 492 EEAGGGEETKSPAEAAASP---EKEAKSPVKEEAKSPAEAK-----SPEKEE 537
 QY 745 AQIFEPHKSIDPSLNEHLIPVYIGHIALADQAFAPKPSWAVPTIVADLMDNRLP 804
 Db 538 AK-----SPEVKSPEKA-----KSP 553
 QY 805 GKTKTKLVPPDEEVSPTMVKQAIKMMVRLMKKNHNSGSTLRLTTIILSDGL 864
 Db 554 AKEEK--SPEAKSPE--KEA-----KSPAEV 578
 QY 865 TEOGKISKPDMSRLAAGSAIVKLAQPCYHEITTEYOYLCALAINDECYOVRQYFAQ 924
 Db 579 KSPKAKSPAKEAKSPAKEAKSPAKEAKSPAEV---AKSPVKEEAKSPAEV--- 632
 QY 925 KLHGLSLRLPLRYMALCALCADPVKERRAHARQCLVKNINVRRETLKHAHVAISELL 984
 Db 633 ---KSPKAKSPKEEAKSPKAKSPEKEE-----AKSPKAKSPVKAKEAKSPKAK 681
 QY 985 SLLEPYVVPYTIHLLHPDVVVKVODIROLKDVKECLMFVLEIMAKNNNSHAFFIRKAV 1044
 Db 682 S-----PVKAEKSPKAKSPVKEEAKSPKAKSPVKE--EAKSPKAKSPVKEEA 730
 QY 1045 ENIKQTK-----DAQGPDAKMKNEKLYTVCDVANNIIMSKSTYVLSLEPKDPLPARFT 1099
 Db 731 KTPKAKSPVKEEAKSPKAKSPKAKT-----LDVKSPEAKTPAKEEARSPPA----- 778

```

OY 1100 QPDKNFSNFKNYLPPMKSFTEPGKPTTNVIGAVNKPSSAGKOSQTKSRMETVSNAS 1159
DB 779 --DKPEFKAKSPKKEVKS---PEKAKS-----PLKADAKAPEKEIPKKEVYK--S 822
OY 1160 SSNNPSPGRIKRLDSSSEMDHSENDYMSPLPKKSKSRDSDLYRELKPKGRK 1219
DB 823 PVKEEEKPOBVKV---EPPKAEKEKAPATPTEKKDSKKE-----EAP--KKE 868
OY 1220 TPVTEOEKLGMDLTKLVQEOKPKGSORSRKGHTASE---SDEQOMPEERKLEKEDIL 1275
DB 869 APRKPYEEK-----KEPAVEKPKESKVAKKEEEDKKVPTPEKAPAKVEKVEDAK 921
OY 1276 ENDEEO--NSPPKKGKRGRRPKPLGGTPEKPEPTMTKSKKSGKSGPPAPEEEBERO 1333
DB 922 PKEKTEVAKKEPDDAKAKEPSKP-----AAKKEAPEKKDKTEEKAK 963
OY 1334 SGTEBKSQSKQHRVRRARQRAESPSSAIESTOSTPOKGRGRPSKTPSPQPK 1388
DB 964 KPEPKRTEKAKEDDKTUSKEPSKPAKEKSSSTDQKSKPPEKATEDKAK 1018

RESULT 9
USOL YEAST STANDARD; PRT; 1790 AA.
AC P2536;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YD058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A.
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL T Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YEL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X54378; CAA38253.1; -
CC EMBL: L03188; AAB00143.1; -
CC EMBL: U53668; AAB66659.1; -

```

```

DR SGD; S0002216; USOL.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR006955; Usol_p115_C.
DR InterPro: IPR006953; Usol_p115_head.
DR Pfam: PF04871; Usol_p115_C_1.
DR Pfam: PF04869; Usol_p115_head_1.
DR KEGG: K04871; Usol_p115_head_1.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (PORENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASF/GIU-RICH (ACIDIC).
FT CONFLICT 847 847 G->E (IN REF. 2).
FT CONFLICT 924 924 E->K (IN REF. 2).
FT CONFLICT 1253 1253 V->I (IN REF. 2).
FT CONFLICT 1319 1319 I->V (IN REF. 2).
FT CONFLICT 1461 1461 N->S (IN REF. 2).
FT CONFLICT 1581 1581 G->S (IN REF. 2).
FT CONFLICT 1600 1600 I->V (IN REF. 2).
FT CONFLICT 1661 1661 R->S (IN REF. 2).
FT CONFLICT 1772 1772 D->DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 3.0%; Score 217; DB 1; Length 1790;
Best Local Similarity 17.5%; Pred. No. 0.016;
Matches 256; Conservative 252; Mismatches 520; Indels 433; Gaps 61;

OY 36 KMYVKTQMDQDQSEKEKELYLMALHLASDFLKHGKDVRLVACGLADIPRIYAPPA 95
DB 599 KSVIKSLISFYIODE-----DVIKLYMLGLVAFESKES 638
OY 96 PYTSPDKLQDIEMFTROL-----KGLDTPKSPQ- 124
DB 639 PEPFR-----KEYFEFIFITLKDNVYASRIKQFKKDSYFSKVDNEDSILNPELDETLG 694
OY 125 -ENRYFYLL--ENIAVWKS-----YNCF--ELEDNSNIFQVLRITLFSVYNNGH 169
DB 695 YFSYFYFLQENENYRIRLRALSHDPEEPISKISFEVEKLOQCKTLGGEITSL--QTE 752
OY 170 NOKVHHMVDLMSIICGDPVSQE-----LDFVLVNLVPAHKLMLNQ 213
DB 753 TESTHENTLTKLALNENHEKDEKQYILNSSHSLKENSILFETELKNVRSLDKMT-Q 811
OY 214 AYDLAKALKRTAQAEPTTFPNQ-----VLMGKTS--ISDLSHEVFD 257
DB 812 LRVLVETKQENQTALETKYKISTIKQEDSITKEGLETLSQKKAEDINMKGRDLFA 871
OY 258 LIEELVNIIDSHLLSVLPQLEFKLSGNDNERLOVYKLAKMGARDSLASQNKPL--- 314
DB 872 LSREMOAVERN-----CKNLQ--REKDSNVNHOKETKSLKEDIYAAKITEKALINENLEEM 925
OY 315 -WQCYLGRENDIHVPIRLCEVFSACHLMNHPDLAKDLTEYLVKVRSHDEE-AIRHDVIV 372
DB 926 KICQ--NNLSKEKEHISKELVEYKSR--FQSHDLVAKLTPEKLSLANNKQMDKAENESLI 982
OY 373 SIYTAKKDILLVNDHL-----NFERER--TLDKRWYRKEMAGLAQIYKKA 420
DB 983 KAVEEKNESIOSLNQNKIDMSOEKENFQIERGSIENIKNIOLKTTISDLEQTEKEII 1042
OY 421 LOSAARKDA-AKQIANTIKKLLHIYVONSIDDLVLRIFAQVMVHNLETTERRMCLVY 479
DB 1043 SKSDSSKDEYESOISLKEK-----LETA----- 1066
OY 480 LVYTLDLNAKALNEMWKNLRLHGVKDLIDIKQPKTDSVAKFAFSKVMVITRNLDP 539
DB 1067 --TTADENVNKISLTK-----TRRELEALAYKNLKNLTK----- 1104
OY 540 GKADFQKFTQVLEDEKIRKQLEVLVSPGSCQAEGCVREITKLGKPKQPTNPFLE 599
DB 1105 -----LETSEKALKEV-----KENEHLKREKIQLEKEARETQKQIN 1141
OY 600 MIFLLERLAPVUID-----TESTSALIKQVKNYSIDSTADDEDEGCVPTDQAIRAGLE 651

```



```

Db 1142 SLRANLESLEKEHEDLAAQLKKEEQIANKEHQYNEEISQULNDEITSTQOENESJK----- 1197
OY 652 LKVLSTFTHPISTHSAEFESLACLKMDDEKVAALAIQFNKTSKLEEDPPIHSALL 711
Db 1198 -----KKNDE--LEGEVAKAMST-----SEQSNKKKSEI 1225
OY 712 PVLHHSKKGPPRQAQVAIHCIHAIFSSKETOFAQIFPLHSLDPSNLEHITPLVITG 771
Db 1226 DALNMQIKE-----LKKNETNEASLESTI KSVSEIVK----- 1259
OY 772 HIALLPDQFAAPKSWATFTIVKDL-MNDRL-----PGKTTKLWPDDEEVSPTWVK 825
Db 1260 -IKELODE-----CNFEKEVSELEDKLKASEDKNSKYLELOKESEKIKRELDA 1308
OY 826 IOATKMWRLMLGKNNHNSKSGTSLRLITTLHSDGLTEQGRKISKDMSRLRLAASA 885
Db 1309 TTEKLITQELKTTNLSKAKESSELSRLKKT-----SSBERKAEQOLELTK----- 1355
OY 886 IVKLAQEPYHEI-TTLEQYOLCALAINDECYQVQVPAOK-----LHKGSLRLPLEYMA 941
Db 1356 -----NEIQKNQAFERKRLNBSSTIQEYSEKINTLEDELIRLQNE-- 1402
OY 942 ICALCANDPVKERBAHARQCLVKNINVRREYLKQHAASEKLLSLPEYVVPYTHLLAH 1001
Db 1403 ---LKAKE-IDNTRSELEKVSLSNDELLEE--KQNTIKS-----LODEILSYKDKITRN 1450
OY 1002 DPDVVKVO-----DIEOLKDYKECLMFVLELIMAKNENNSIAFLIKKAVENIKQTKDAQCP 1056
Db 1451 DEKLISTERDNKRDLSESLK-----EQLRRAQESKA-----KVEEGKLKLEESSEK 1495
OY 1057 DDAKANEKLYTVCDVAMNIIKSKSTTYELSEPKDVLPAFPOTOPDKNFSMTKNVLPPEM 1116
Db 1496 EKAL-EKSKEMMKKLESTISNET--ELKSMETI-----RKSDLEBOSKSAEDTI 1546
OY 1117 KSFTPGKPTTNVLAGVN-----KPISSAGKOSQTSKSMETVSNASSS-----NP 1164
Db 1547 KNL-----QHEKSDLSIRINESEKDIIEELKSKLRIFAKSGSELETYKQELNNAQEKIRINA 1602
OY 1165 SSPQRIGRLDSSSEMDHSENDYTMSSP-----LPGKSDKRDSDLYRSELEKPRGKK 1219
Db 1603 EENTVLKSLKLEDIRELKDQAEIKSNOEKEKLLTSRLKELEQELDSTQQAOKSSEBER 1662
OY 1220 TPYTE-OEEKIGMDLTKLVQ-----EQKPKGSGRSRRKRGHTASEDEQOWPEKKR 1269
Db 1663 AEVARKFYEQSQLODEKAMLETXYNDLVNKRQAMKRDDDYAK-----TTTSQKQELTKL 1717
OY 1270 LKEDILENEDQNSPPKKGKRGPRPKPLGGTPKPEEPTW-----KTSKKGSKKK-- 1318
Db 1718 AKE--LDNLKAKENSKLKEANEDR-----SEIDMLMLVTDLDEKNAKRYRSKLKDLG 1766
OY 1319 -----SGPPAPEEEEEEROSG 1335
Db 1767 VEISSDEBDEDEDEDEEERG 1787

```

```

RA Starr D.A., Han M.;
RT "Role of ANC-1 in tethering nuclei to the actin cytoskeleton.";
RL Science 298:406-409(2002).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN-Bristol N2;
RA Gattung S., Goela D., Brooy M.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 689-1243 FROM N.A.
RX MEDLINE-22296983; PubMed-12408964;
RT Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT "The nesprins are giant actin-binding proteins, orthologous to
RT Drosophila melanogaster muscle protein MSP-300.";
RL Genomics 80:473-481(2002).
CC -1- FUNCTION: Plays a central role in nuclear and mitochondrial
CC anchoring. Probably connects nuclei to the cytoskeleton by
CC interacting with unc-84 at the nuclear envelope and with F-actin
CC in the cytoplasm, creating a bridge across the nuclear envelope
CC between the cytoskeleton and the nucleus.
CC -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.
CC Interacts indirectly with unc-84.
CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (potential). The
CC largest part of the protein is cytoplasmic, while its C-terminal
CC part is associated either with the nuclear envelope, most probably
CC the outer nuclear membrane, or with mitochondrial membrane.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in all postembryonic
CC cells.
CC -1- DEVELOPMENTAL STAGE: First expressed in L1 larvae, and thereafter.
CC -1- DOMAIN: The large coiled coil domains are composed of 6 nearly
CC exact repeats of 903 residues. The last repeat is partial. These
CC repeats are conserved in Hawai (CB4856), Australia (AB4) and
CC Germany (RC301) strains. The length of the repeat may be
CC maintained because of a selective advantage to keep the protein
CC large and allow a single molecule to extent more than 0.5
CC micrometers.
CC -1- DOMAIN: The klarsicht domain, which contains a potential
CC transmembrane domain, is essential for the nuclear envelope
CC targeting.
CC -1- SIMILARITY: Belongs to the Nesprin family.
CC -1- SIMILARITY: Contains 1 actin-binding domain.
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -1- SIMILARITY: Contains 1 klarsicht domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AY157938; AAN35200.1; -
DR EMBL, AC006834; AAF40010.3; -
DR EMBL, AY126454; AAM95163.1; -
DR HSSP, P46939; 10AG.
DR WormPep: ZK973.6; CE33588.
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF00307; CH; 2.
DR SMART: SM00264; BAG; 6.
DR SMART: SM00033; CH; 2.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE: PS50021; CH; 2.
DR Structural Protein: Cytoskeleton; Actin-binding; Transmembrane;
DR Coiled coil; Repeat.
FT DOMAIN 1 8494 CYTOPLASMIC (POTENTIAL).

```

FT	TRANSMEM	8495	8513	ANCHOR FOR TYPE IV MEMBRANE PROTEIN (POTENTIAL).
FT	DOMAIN	8514	8545	PERINUCLEAR SPACE (POTENTIAL).
FT	DOMAIN	1	325	ACTIN-BINDING.
FT	DOMAIN	23	130	CH 1.
FT	DOMAIN	222	325	CH 2.
FT	DOMAIN	3241	8199	6 X TANDEM REPEAT.
FT	REPEAT	3241	4143	1.
FT	REPEAT	4144	5097	2.
FT	REPEAT	5098	6000	3.
FT	REPEAT	6001	6903	4.
FT	REPEAT	6904	7806	5.
FT	REPEAT	7807	8199	6.
FT	DOMAIN	754	774	COILED COIL (POTENTIAL).
FT	DOMAIN	1072	1101	COILED COIL (POTENTIAL).
FT	DOMAIN	1215	1236	COILED COIL (POTENTIAL).
FT	DOMAIN	1324	1384	COILED COIL (POTENTIAL).
FT	DOMAIN	1574	1629	COILED COIL (POTENTIAL).
FT	DOMAIN	1725	1754	COILED COIL (POTENTIAL).
FT	DOMAIN	1950	1981	COILED COIL (POTENTIAL).
FT	DOMAIN	2103	2580	COILED COIL (POTENTIAL).
FT	DOMAIN	2682	2712	COILED COIL (POTENTIAL).
FT	DOMAIN	2852	2949	COILED COIL (POTENTIAL).
FT	DOMAIN	3002	3119	COILED COIL (POTENTIAL).
FT	DOMAIN	3178	3295	COILED COIL (POTENTIAL).
FT	DOMAIN	3346	3417	COILED COIL (POTENTIAL).
FT	DOMAIN	3482	3552	COILED COIL (POTENTIAL).
FT	DOMAIN	3587	3703	COILED COIL (POTENTIAL).
FT	DOMAIN	3781	3839	COILED COIL (POTENTIAL).
FT	DOMAIN	3902	4022	COILED COIL (POTENTIAL).
FT	DOMAIN	4114	4198	COILED COIL (POTENTIAL).
FT	DOMAIN	4249	4320	COILED COIL (POTENTIAL).
FT	DOMAIN	4436	4506	COILED COIL (POTENTIAL).
FT	DOMAIN	4541	4657	COILED COIL (POTENTIAL).
FT	DOMAIN	4735	4793	COILED COIL (POTENTIAL).
FT	DOMAIN	4856	4976	COILED COIL (POTENTIAL).
FT	DOMAIN	5035	5152	COILED COIL (POTENTIAL).
FT	DOMAIN	5203	5274	COILED COIL (POTENTIAL).
FT	DOMAIN	5339	5409	COILED COIL (POTENTIAL).
FT	DOMAIN	5444	5560	COILED COIL (POTENTIAL).
FT	DOMAIN	5638	5696	COILED COIL (POTENTIAL).
FT	DOMAIN	5759	5879	COILED COIL (POTENTIAL).
FT	DOMAIN	5938	6055	COILED COIL (POTENTIAL).
FT	DOMAIN	6106	6177	COILED COIL (POTENTIAL).
FT	DOMAIN	6242	6312	COILED COIL (POTENTIAL).
FT	DOMAIN	6347	6463	COILED COIL (POTENTIAL).
FT	DOMAIN	6541	6599	COILED COIL (POTENTIAL).
FT	DOMAIN	6662	6782	COILED COIL (POTENTIAL).
FT	DOMAIN	6841	6958	COILED COIL (POTENTIAL).
FT	DOMAIN	7009	7080	COILED COIL (POTENTIAL).
FT	DOMAIN	7145	7215	COILED COIL (POTENTIAL).
FT	DOMAIN	7250	7366	COILED COIL (POTENTIAL).
FT	DOMAIN	7444	7502	COILED COIL (POTENTIAL).
FT	DOMAIN	7565	7685	COILED COIL (POTENTIAL).
FT	DOMAIN	7744	7861	COILED COIL (POTENTIAL).
FT	DOMAIN	7912	7983	COILED COIL (POTENTIAL).
FT	DOMAIN	8048	8118	COILED COIL (POTENTIAL).
FT	DOMAIN	8153	8204	COILED COIL (POTENTIAL).
FT	DOMAIN	8273	8329	COILED COIL (POTENTIAL).
FT	DOMAIN	8370	8390	COILED COIL (POTENTIAL).
FT	DOMAIN	8486	8545	KLARSICHT.
FT	DOMAIN	1655	1658	POLY-ARG.
FT	DOMAIN	3021	3029	POLY-LYS.
FT	DOMAIN	3924	3932	POLY-LYS.
FT	DOMAIN	4878	4886	POLY-LYS.
FT	DOMAIN	5781	5789	POLY-LYS.
FT	DOMAIN	6684	6692	POLY-LYS.
FT	DOMAIN	7587	7595	POLY-LYS.
SQ	SEQUENCE	8545 AA;	956470 MM;	02A94D94BEE19E3 CRC64;
Query Match 3.0%; Score 212.5; DB 1; Length 8545;				
Best Local Similarity 18.4%; Pred. No. 0.2; Mismatches 642; Indels 495; Gaps 79;				
Matches 317; Conservative 268;				

QY	19	KKEISDKIS-KKEWVRRLKMYVKTFMD-----	MDQSEEREKELYNLALHL-----	ASDF 67	
DB	5221	VNEIAPKLDLVQSQLOSVQEPAPASLDQKQLEEDENQKNLENLNLNLPENDPTADEL		5280	
QY	68	FLKHFG-----	KDVLIVACCLADIRIYAPAPYTSPPKLDIFMTITROKLGEDTYS	122	
DB	5281	ROKSQWDLRLKDLQOLGSAGVD--	KLAALAAFNAAKNMADALDITRE--	DGGDNKS 5337	
QY	123	P-----	QENRYEYLENLAWKSYNICEPELDS-----	NEIPTOLYRLFYSYNGK 168	
DB	5338	PDELIDDLAKKEETVAKLDITVYSGVR----	DELDDEKRAEYNDLARL-ATAADVLKN-	5391	
QY	169	HNQVHMHVMDLMSIICGDTVS--	OELDDITVLVNLVPAKMLNKQAYDLKAL-----	221	
DB	5392	-----	KRAELQAVAKADEKSLHSDVDIRIYSRLVYRESDELRRH-AEAVPTQYAP	5443	
QY	222	-----	LKRTQAALIEPTITTF-----	FNQVIMLKTSISDSSEHVFDLILELYN----	264
DB	5444	KAEELKEVEAAKAAVIANAPSSDAHVOLEQAVATAETLIPLEERA-----	RLMNEFLA	5498	
QY	265	-----	IDSHLLSVLPQLEFKLS-----	NDNEERLOVVKLL--AKMFGAKDSEIASON	311
DB	5499	ARNDDID-----	ALIBOLOQPLDAVLNAQPKRSABEAAQDVENLRNNSQOQLSDLDKNTANQ		5553
QY	312	K-----	PLMQCYLG-RF-----	NDIHVPIRECV--KFASHCLMHPD	346
DB	5554	RISELDDPLESAVADVFEDVDACQTRHQYDVLNDVNAELEDEFTLKQASQVANEID		5613	
QY	347	LAKDLELYKVRHDEEAEARHVIYSIYTAAKKDLLVNDHLNFRVR----	TLQKRRR	403	
DB	5614	ISK-----	MIDSTDERSTLDRTAKSDIPALAAQINRIKDRIVNADASKRHVTYDPKTA		5667
QY	404	VRKEAMGIAQIYKKKALOSAGDKAOKIAMIKIDLHIYYONSI-----	DBRLAYE	456	
DB	5668	EDLDNKLAKQTLDAIKITSDEHKEQDLISLK--	LNISQEQIPLDQKSDDLKTRAE		5724
QY	457	RIFAQYVPHNLEETTERMKCLYYLVATFLDINAVKALNEMKCONLLRHQVKDL-----		510	
DB	5725	KEITNSLKPEAEPL-----	LAKIOELREAKRVGDEARSAAHQDVALERE		5770
QY	511	-----	DLIQPKP--	TDAVSALFESKVM-----	ITRNLPRDGKAQDF 545
DB	5771	AEDVTAKESAKKKKKKKSSPQEMIDELSAKVVEAKALIPKIEBAKNEPLP-----	ADD	5825	
QY	546	MKKEFTVLEDEDEKIRKOLEVLVSPSCSKQAE-----	GCVEITKRGKNPKQ-----	PT 594	
DB	5826	KPKAEQVLSNLEAFVQVDETQVSE-----	KODELDKLNANDAIKRIKGDALDADEKTVPS		5881
QY	595	N-PLEMIKIFLERIAPVHIDT-----	BSISALIKQVNSIDGTA		633
DB	5882	SVPALESEK--	DRIAP-HIATLVEAVNDVPASVEPSAVALRRAKAFVSDEKKNIOQTG		5937
QY	634	DDE-----	DEGVPTDA-----	IRAGLELKVLSTHRI	662
DB	5938	DDEKRADELKNDGNAVKNVEDVSKYQNPQPLDVAKDANKLKTATVQDLKLA-----		5992	
QY	663	SFSAETPESLACLKMDKVEALAO-----	IFKNGSKIEEDFPHRSALLP		712
DB	5993	--	ESSKIDPQVAKDKDSTKAKELLQALEKAIPOEDAIRQOALINRNLNLEKELTK		6050
QY	713	VLHHKSKGPPROAKYAIHCIAHIFSSKETQFAQIEPLRLKSLDPSNLEHLITPLVIGH		772	
DB	6051	VDEFKPED-----	ALPIVDQLANITNL-----	KRTDSSNNEKAVAPSSLSH	6093
QY	773	IALLADQFAAPKSNVAFIYVD-----	LMDNDRLPGRKTTKLVMPDEE--	VSPB 821	
DB	6094	DDLIVV-----	GLPEKVPOLAHADKQALNKAAVNEIAP-----	KLOVSOQOLOSVQEP 6144	
QY	822	TWVKIQAIKMWVWMLGKNNHNSKSGSTLRLITTLHSGDLEEGCKISKIPMSRLR--		879	
DB	6145	VPSLDEQOKQLEDEVNOKHNLE-----	NLIANLPENDPTADELROKSQWDLRLKDL		6197

```

OY 880 -----LAAGSAIVKLAQPCY-----HEITLEQYCALAI 911
DB 6198 LKQGSAGVDKIALAALAAFNAAKNADALDITREDGGDNKSPBELI----- 6245
OY 912 NDECYQVROVPAQKILH--KGLSRLRP---LEYMAICA--LCARDPVKERRAHQCIY 963
DB 6246 -DLAKKEETVAKLIDTVSGVPPDELDDKERAEYNDLLARLATADVLKKNRAELQEA-- 6302
OY 964 KNINVRREYIKQHAANVSEKLSLIP-----EYV-----VPYIY 997
DB 6303 --VAKADEKSLHDSVDRIVSRLVPSRESELRHNAEVPQYAPKAELKEVEAKA 6360
OY 998 LLADHPDY-VKYOIDEQ-----LKDYKEGLMFWLETLAKKNENNSHAFIRKVENIX 1048
DB 6361 VIANPSSDAHVOIQAVATAETIIPDLERARLWNEFLARND-----IDALIEQ 6414
OY 1049 QTKDA-----QGPDDAKKNEKLYTVCDAVMNIMSKSTYSLESPKDPVLPARF 1097
DB 6415 QPLDAVLAQPKRSAAEAQDVENLNNSQQLSDLNKIANQRISELDPLESAVADRF 6474
OY 1098 F-----TOPDKNFSNTKNTLPPEM--KSEFTPEKPTTVNLCAVNRPLSSAGQSQTKSSR 1151
DB 6475 FDVAEQTROHQYDVLNDVALEDEETLLKOSASQVANEIDISKMDS---TDPERSI 6530
OY 1152 MEVSNASSSSNPSSPGRIGRL-----DSEMDHSENEEDYTMSSPLPG--KK 1197
DB 6531 LDTAKSDIPALKQAINIKRIVADASRKHVTTDPKIAEDLDKAKLQIETLDAIKT 6590
OY 1198 SDRKDDSLVSELEKPRGKKTPTVEQ---EKLGMDDT---TKLVQEQPKGSG-- 1247
DB 6591 SDEHKEQILIS-----LKLNISQFEQIPLDQDKLDTAKKEITNSLKPEAEPL 6642
OY 1248 -----RSRR-GHTASESDQEQWPEEKRLKEEDLENEDEQNSPPKGRPRPKPLG 1299
DB 6643 LAKIOELREARVGDARSAAHDQVLALEKEADVTAKESAKK--KKDKKSPQEMIDE 6700
OY 1300 GTPK--EETPKTKSKGSKKSGPAPPEEEERQSGNTEQSKSKOHVSRRAQOARF 1357
DB 6701 LSAKVVEAKALIPKEEAAKNENLPADDKPKAEQVLS-NLEAFVADVEQVSEKQDELDK 6759
OY 1358 SPSS-----AIESQSTPPQKGRPKTPSPSQPKKNV 1391
DB 6760 LNNANDAIKRLGDLDAEKTIV-----PSSVPALSEKDR 6796

```

```

RESULT 11
NFH_RAT STANDARD: PRT: 831 AA.
ID NFH_RAT P16884; Q63368;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).
GN NEFH OR NEFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=3143606;
RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites.";
RL FEBS Lett. 241:213-218(1988).
RN [2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Phan-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and

```

```

RT in situ detection.";
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN [3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderson B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H).";
RL FEBS Lett. 209:203-205(1986).
RN [4]
RP SEQUENCE OF 318-831 FROM N.A.
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spiner N., Snyder S., Anderson J., Goldhaber D.,
RA Smulowitz M., Carroll Z., Emanuel B.S., Bretner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37227; AAAA1693.1; ALT_FRAME.
DR EMBL: X13804; CA932038.1; ALT_FRAME.
DR EMBL: M21964; AAAA1695.1; -.
DR EMBL: J04517; AAAA1692.1; -.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; phosphorylation; Repeat.
FT NON_TER 1
FT DOMAIN 1
FT 276 641 51 x 3 AA TANDEN REPEATS OF K-S-P.
FT CONFLICT 164 164 L->I (IN REF. 2).
FT CONFLICT 185 185 L->S (IN REF. 2).
FT CONFLICT 193 193 L->T (IN REF. 2).
FT CONFLICT 199 199 M->T (IN REF. 2).
FT CONFLICT 346 346 K->N (IN REF. 1).
FT CONFLICT 373 373 A->V (IN REF. 1 AND 4).
FT CONFLICT 482 482 G->E (IN REF. 2 AND 4).
FT CONFLICT 485 485 P->S (IN REF. 2).
FT CONFLICT 570 571 RK->KE (IN REF. 2 AND 4).
FT CONFLICT 591 591 P->T (IN REF. 2 AND 4).
FT CONFLICT 727 727 A->V (IN REF. 4).
FT CONFLICT 757 759 AAP->GST (IN REF. 4).
FT CONFLICT 769 769 T->L (IN REF. 2).
FT CONFLICT 775 775 R->P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3f13EF768 CRC64;

```

```

Query Match 2.9%; Score 212; DB 1; Length 831;
Best Local Similarity 22.9%; Pred. No. 0.01;
Matches 85; Conservative 56; Mismatches 150; Indels 80; Gaps 13;

```



```

OY 445 -YONSI-----DDRLVERIFAQVWVPHNLETTERMKLYIYATL-----DLNAVY 490
DB 4163 DIQSALSTVSTOAEADSRLLTE-----ISSWDGCKYFYRSIARRARLMAAL 4210
OY 491 A-----LNEAMKCONLLRHQVLDLILKOPKTDASKAJESKVMVTRNL-----536
DB 4211 ATPAKEMGMGNVERGRFSAHLMK---MLYRQRBSITTL-----SEOWILIRNLSCVOET 4263
OY 537 -----PDPGAODEMFKFTYOVEDEDEKIRKQLEVVS-----PTCSCKQAE 577
DB 4264 HSRLLMGPOAYPAVAFPPDQVQWTERL---QHLAMQOQLLBQLSLMLCCSVGAPRGH 4320
OY 578 GCVRITTKKLGKPKOP-----TNPLEMIKFLERIAVY-----611
DB 4321 GNV-----OVLGQPPGPELSEGPLSKGOLCGVLDLPSNLSTYSPSPGSQLPSCGRMRKO 4376
OY 612 -HIDTESIALIKQVNSIDGTADEDEGVPTDQAIRAGLELLKVLSTFHPISFSAETP 670
DB 4377 DHLMOGSTTRL-----TEMLKTIKIVKADVD---KIRQSCETLPHSMKPE 4419
OY 671 E---SLACL-----KMDKVAEALQIFKNTGSKIEEDPH 705
DB 4420 EVCSSALSLQSVYHLOGLSLFTLPGMEVEQOROSOMALVLEVRGEISKAMADFTT 4479
OY 706 IRSALL-----PVLHHSKSKGPPROAKVAIHICHAIFSSKEQFOAIFEPKHSIDP 757
DB 4480 KTHLLTSDSGONGMDEGEVEDSEQMEIARILCAIQVLEEKKNKAEEINTQASP 4539
OY 758 ---SNLEHLITPLVTIGHILALAPDOFAAPKSWAVAFIVDMLMNDRLPGKTKTKLV 813
DB 4540 QEDYAGFERLQS-----GHILFKLEBDDF-----WA 4564
OY 814 PDEEVSPTMYKIQAIKMMVRLGLMKNHKSSTYLRL-----LTTILHSDG 862
DB 4565 ---DVS---TLHVQITISAISELLERLKSXYGEDTAKLFPQSOSCLVRLVPLYSYS 4618
OY 863 DLTEQGISKPDMSRLRAGSAIYLAQEPCEYHITLEVOQLCALIINDCYQVROFE 922
DB 4619 DLV-----LEFLTMSLATRSTAKL-----LSVLA-----QVF 4646
OY 923 AOKLHKGSLRLRPLEYNAICALKADPKYERRARRAOLVKNINVRREYLKOHAAVSEK 982
DB 4647 TELAQKGRFC---LPREFM-----EDSAGEGATE-----4671
OY 983 LLSLLEPYVPTTILHLAHDPDYVYQDIEOLKDYKECMLFVLEILMAKNENNSHAFTRK 1042
DB 4672 -----FHDEGGGIGEGEGMKDVSD-----QIGNEQOVEDTQK 4705
OY 1043 MYENIKOTKDA---QGRPDA-KMNE---KLITYCDVAMNIMSSTTY-----S 1084
DB 4706 GOEKDEKPDKSDIKGEDNALIKESDEDDGKMHODELEBOEEDDEKSDSEGDLKMGD 4765
OY 1085 LESPRDVLPAFRFTQDPKNSNTKNYLPRPKKSFTT-PGKRTTNVLGAVKPLPSAGK 1143
DB 4766 LNGEADKLDERLMDDEDEDEDE---DNKTETGQMGDEEDSELVAKDNLDSD-CN 4820
OY 1144 OSQRTS---SRMETVSNASSSSNPSPGRIKGLDSEMDHSHENEDYTMS---PLPGKKS 1198
DB 4821 SNKDSQODKKEEKEAEADOGGOGEDKINEIDRIDRYDENEDVDYHGQOEKVPRP-EAL 4879
OY 1199 DKRDSDLVSELEKPRGRKKTPTVYEOEELGLMDLTLKYOQKSGORSKRKHT---1255
DB 4880 DLPRDLNL---DSEKKNGEOTDNEGEENPLE-----IKKPEAGHAEAEKEETED 4931
OY 1256 ASESDQOPEEKRLKEDILENEDEON---SPPKKGRGRPRPKPLGSGTPKEPTTKTSK 1313
DB 4932 QNESQSPQEPGPESEDOKAGEEEMDTGADQDQDAQNHPE---HSPEEQOQSVEEKDK 4988
OY 1314 GSKKSGPPAP-----EEEEEROSGNTBOK---SKSKHRSRRRQARSPSS-A 1363
DB 4989 EADEGGENGADGQFOPEEEREEDSDTEEOVPALEBRKEH---ASCQGTGVENMQNTQA 5046

```

```

OY 1364 IESTOSTPOKGRGR 1377
DB 5047 MELAGAAPEKEQK 5060

RESULT 13
BPAL_MOUSE
ID BPAL_MOUSE STANDARD: PRT: 7389 AA.
AC Q91206; Q91207;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1, isoforms 1/2/3/4 (BPA) (Hemidesmosomal plaque protein) (dystonia musculorum protein) (dystonin).
GN BPA1 OR DST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RC STRAIN-BALB/c; TISSUE=Muscle, and Neuron;
RX MEDLINE=21405767; PubMed=11514586;
RA Leung C.L., Zheng M., Prater S.M., Liem R.K.H.;
RT "The BPA1 locus: alternative splicing produces multiple isoforms with distinct cytoskeletal linker domains, including predominant isoforms in neurons and muscles."
RL J. Cell Biol. 154:691-697(2001).
RN [2]
RP SEQUENCE OF 6693-7389 FROM N.A. (ISOFORMS 3 AND 4).
RC STRAIN-C57BL/6J; TISSUE=Fetal skin, and Fetal spinal cord;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikiido I., Osato N., Saito R., Suzuki H., Yamanaoka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schiml L.M., Kanaplan A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Gilmour S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedziński R.M., King B.L.,
RA Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltakis L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelina A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilting L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carlini P., Hayatsu N.,
RA Hironaka-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Atakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [3]
RP FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing intermediate filaments to the inner plaque of hemidesmosomes. The proteins may self-aggregate to form filaments or a two-dimensional mesh (By similarity).
CC -! SUBUNIT: Homodimer. Interacts with the neuronal intermediate filament protein, Prph (By similarity).
CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -! ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=7;
CC Name=2; Synonyms=b;
CC IsoId=Q91206-1; Sequence=Displayed;
CC Name=1; Synonyms=a;

```



```

OY 126 NRYFYLLENIAWVKNICFELEDSENLFTOLYRFLSYINNGHOKYMHVMDLMSIT 185
DB 4802 SAKLDVLE-----SLNSOKDGFKFTTEOSNITEKTSISEG-----LL 4840
OY 186 CEGDVTSELDLTVLVNIPPAHKNLNKOAYD-----LAKAL-LKRTQAIEPYIT-- 234
DB 4841 LKTOGAERKALOLOJNTKKTDMDRRKQYKEREKLSOKALKYREVETLRWIDRC 4900
OY 235 -----TFPNOVLMGKTSISDLS-----EHVFDLIELYNIDSHLLSVLPOLFELK 282
DB 4901 OHSLDGVTFSLDPTSESSIAELKSLQKEMDHFQGM-LELLNNTANSLSLVC-EYDKAV 4958
OY 283 SUNDNEERLOYVLLKMGAKSELSAOKRPLMOCY-----LG 320
DB 4959 TEENSEKRYVRYEOLQSKTVSLNNAQKFEQOVSRTQROLODTKEOLEVHNSIG 5018
OY 321 --RFNDIHVRIRLECVKFPASHCLMNHPRDLAKDTEYLVKVRSHDPEALRHDIYSIVPA 378
DB 5019 POAYSNKHLSV-LQAOKSLSOTLKQOVDKARLADLVVEADSKGT--SDVLLQAFETLA 5075
OY 379 KKDILLVND-----HLNFVRE-----RTLDK 400
DB 5076 EHSSELQOQVDEKCSFLETKLQGLGHFOVTLREMSQFTECDELDGMAPVGRDAETLRK 5135
OY 401 R-----WVRKEAM-----GLAQYKKYALQSAAGKDA 430
DB 5136 OKACMOFTLKKLEALMASNDSANRTCKMMLATEETSPDLIGVKKOLEALSKQCNLLDRA 5195
OY 431 R-----QIAMIKDLHLHYQ-----NSIDRLIVERIF-- 459
DB 5196 KTRBOVGATGEKLEEFHRLKEEFSTLQKAEHESOGPVGTETETINQOLDVKKVQK 5255
OY 460 -----AOVMVPHNLEETBRMKCLYLYATLDLNAVKALNEMK----- 497
DB 5256 EIEPILOYKQODVNMWLGGLIQSAANNTQGLEHDL-----SVNSRKTLNKKVQAORT 5310
OY 498 --CONLRRH--QVKDLDLI-----KPKTQAS-VKAIFSKVAVITRNLPD 538
DB 5311 SOLQALHLHGRFOALLESLSMADTEELVANOQPPSAEFVVAQAQOLEOKLDRLLLED 5370
OY 539 PGKADQFMKFTQVL-----EDDEKIRKOLEVNS-----PTSCQKQAEQCVREIT 584
DB 5371 RKSJVEVIRREGEKIAASAEPADRVKLTROLSLDRMBALLSRAEARRQOEG-ISVYA 5429
OY 585 KKLGRKOPTNPFLEMIKIFLERLAPVHIDTESIALKQVKNKSIDGADDED--EGVP 641
DB 5430 QEFHGTLPLEMLVAVEKKLANSEPIGTOAPKLEEOISQ-HKAL-----QEDILRKQS 5483
OY 642 TDOAIRAGLELLK-----VLSFTH-----PISPHSAFTFESLL----- 674
DB 5484 VDQALLNLELLKQTTGDEVLLIOLKLEAIRKYDITKLSADVAKTJHEHALQLAGQLOS 5543
OY 675 -----ACLKMDDEKV-----AEALQI-----FKN--TGSKIEEDPPIRSAL 710
DB 5544 MHKELCNMLDKVEVELSTYEQGLKGAASQVQEQKLEKNVRSKNKALVYSLNVSALL 5603
OY 711 LPLVHNKSKKPPROAKYAIHICIAIESSKETOPQOIEPRLKSLDPSNLEHLITPLVYI 770
DB 5604 LELVPCRAKEG-----LEKTIADDBEPL-PCDEPQSRHKHVEI--- 5641
OY 771 GHIALAPDOF---AAPKSMVATPIVKDLAMD-RLGKGTKTKLWVDEKSPETWMI 826
DB 5642 -DAILRSQOEQADADAELSWTTETOKTMSIGDIRL-----ODQTSAL 5686
OY 827 QAIKMMVWMLGMMKNHNSKSGTSTLRLTTLLHSQDLEEGKISKPRM----- 875
DB 5687 QVQAKFTMDILRHKOIIDELVTSIGKIMT-----SGEEKQSMKKKLDKVLKTYAVQOI 5742
OY 876 ---SRLLAAGSAIYKLAOEPCYHEI---ITLEQ---YOLCALAINDECY-----QV 918
DB 5743 NSERHLOLERAOSIVSQFWE--TYEELMPWLTETQRIISOLBPAPALEYETFERQOEHRQL 5801

```

```

OY 919 ROVFAO-----KLHK-GLSRRL-PLF-----YMAICALA--KDPYKERRAHQOC 961
DB 5802 RELLEHNRPHIDKMKKTPOLLELSPKGIYIOEKTVADTLVSOIKEDVKKRAV----V 5857
OY 962 LVKNINVRREYIKQHAASEKLSLLPYYVVPYTIHLHAHPDYKVQO-----DIE 1012
DB 5858 LDEAISQSTOFHDKIDQILLESLEIRIAERLRQPPS---SABVEKKEQIEGENKSVYME 5914
OY 1013 QLKDVKECLMFVLEITLMAKNEN-----NSHAFIRKVENIKOTKDAQOPDD 1058
DB 5915 KLOPLVETLRQGEEMIRASBGETEKDVSARAVODKIDQVFIWGSIHVLVEERAKLLDV 5974
OY 1059 AKMNEKLYTQDVAMNIMTSSTYLSLESPKDPVLPARFTPOPKNFENFK---NYLPPE 1115
DB 5975 MELAEKEW--CDHMSLVYITQDQFIHDLDEPGIDPSVVKQOQEAABAIEEIDGLOE 6032
OY 1116 MKSEFT-----PGKPRITNVGAVKKPLSSAKOSQOTSSRMETVSNSSSNP 1164
DB 6033 LDMVITLGSSELLAAGCEPDRPIVKKSIDELNSAMDSLKAKKDRVDRLEFAMQAAVQYOD 6092
OY 1165 SSPG-----RIKGRDSSMDHSENDYTK-----SSPLP 1194
DB 6093 GLQGFEDWDIAGDKLATMSPIGIDLETVKQOIELEKQKSEAYQOQEMERLHQAEEL 6152
OY 1195 GKSDKRDSDLVRS---ELEKPRGRKKTPTYEOEKL-----GMDUL----- 1234
DB 6153 LKVTVEADKHTHYQDPLMEKALIMDSIDBERIVRHKLEGTLLALGOFQHALDELAMLT 6212
OY 1235 -TK-LVQDQPKGS-----ORSKRGHTASESDE----- 1261
DB 6213 HTKGILSQKRVGCGPRKAIEIELAKNHVLQNDVLAHQSTVAVNACNDLIESSEGBAS 6272
OY 1262 -----QOMPERKRLKDIILENEBQNSP-----PKKGRGR----- 1292
DB 6273 NLQYKRLILNORW-----ODILEKTDORQOOLDSALROAKGFGEIEDLQWLTDTER 6325
OY 1293 ---PPKPLGG--GTPKEE 1305
DB 6326 HLLASKPILGGUPETAKEQ 6343

```

```

RESULT 14
A3D1_HUMAN
ID A3D1_HUMAN STANDARD: PRT: 1153 AA.
AC Q14617; Q00202; Q75262; Q96G11; Q9H3C6;
DT 28-FEB-2003 (rel. 41, Created)
DT 28-FEB-2003 (rel. 41, Last sequence update)
DT 15-SEP-2003 (rel. 42, Last annotation update)
DE Adapter-related protein complex 3 delta 1 subunit (Delta-adaptin 3)
DE (AP-3 complex delta subunit) (Delta-adaptin).
GN AP3D1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 1
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC TISSUE=Fetal brain;
RC MEDLINE=97447555; PubMed=9303295;
RA Ooi C.E., Moreira J.E., Dell'Angelica E.C., Poy G., Wasserman D.A.,
RA Bonifacio J.S.;
RT "Altered expression of a novel adaptin leads to defective pigment
RT granule biogenesis in the Drosophila eye color mutant garnet.";
RL EMBO J. 16:4508-4518(1997).
RN 12
RP SEQUENCE FROM N.A. (ISOFORM 2), AND TISSUE SPECIFICITY.
RC TISSUE=Heart;
RC MEDLINE=97296491; PubMed=9151686;
RA Simpson F., Feden A.A., Christopoulos L., Robinson M.S.;
RA "Characterization of the adaptor-related protein complex, AP-3.";
RL J. Cell Biol. 137:835-845(1997).
RN 13
RP SEQUENCE FROM N.A. (ISOFORM 1).

```



```

QY 867 --OGKISKPD-----MSRLRLAAGSA-----IVKLAOEPCHYEITILEOY---- 904
DB 363 LLYGWSKKNLMEIYKLTMTVDKAGETTYRDELLTKIDIOSQSYQITNFWMTISL 422
QY 905 -----OLCALINDECYQROY--FAQIKHLKLSRLPLEYMAICALCAKDP 950
DB 423 VELTRLEGRHGLLAOMLDVAIRKAIKRFKFA-----VSOMASLDSANHLA----- 470
QY 951 VKERRAHAROCIVKNINVREYLKQAAVSEKLSILPEYVYPIYTHILLANDPDVYKOD 1010
DB 471 SSTORNGICEVLYAAMWIGEFSEHLQEPHHTLEALRRVTTLLEGHIDA-----VYQVN 525
QY 1011 IEOLADVKECLWFEVLEILMAKNE-----NNSHAFIRKMYENIQOTDAQCPDDAKMKEXLY 1066
DB 526 VVKL-----YASIILOQKQAGAEAGAQAVTOLMDRLQFQVQSA---DLFVQERAS 573
QY 1067 TVCDVAMNIIIMSKSTTYSLESKPDYLPARFPTQDPKNFSNTRKNTLPPEMKSEFTPGKPK 1126
DB 574 CILQILVKHI-----OKLOAKDVPA-----EEVSALFAGELINPVAPKAK--KVYPVE 619
QY 1127 TTNVLGAVYKPLSSAGKOSQTSRMEYVSNASSSNP-----SSPGRIKGRIDSSMD 1180
DB 620 GIDLDMATINEPLS-----DSESDERPRAVFHEEORRPRKHPSEADEELARRRERKOE 675
QY 1181 HSENDYTWSPPLPKGKSDKR---DSDLYRSELEKPRGRKKTPTVEQ-----EKLGM 1232
DB 676 QANNFYIKSSPSPOKRRQDTPGVHEIPVQIDLSVPLKVPGLPSDQVYKLEERRNHQ 735
QY 1233 DITKLVQOKPKGSPRSKRKHTA---SESDEQOMPEEK--RIKEDILEN---EDQNSFP 1285
DB 736 KLEKRRRRKREKREKRRHSLSPTESDEDIAVQOVDYTEEMPEALPSDDDKDP 795
QY 1286 KKGKGRP---PKPLGGGP---KEEPMTKTSKSKS-----KSGCPARPEE---EEBE 1331
DB 796 NDPYALALIDDKPLADSEKPLIQKHRTETSKSPKQDVYKESKPKKKEKKHKE 855
QY 1332 ROSGNTQSKSKQHRVSRRAQRAESPESASIESTQSP---QKGRGSPKTPSPSQ 1386
DB 856 RDKKKKKKKKKSPKPKKKKRRKKEERTKCKKSKKOPPOSEEAGAPVQNGADEE 913

```

RESULT 15

```

ID NP14_HUMAN STANDARD: PRT; 699 AA.
AC Q14978; O15030;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nop140) (Nucleolar and colled-body
DE phosphoprotein 1).
GN NOLC1 OR KIAA0035.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Leukemia;
RX MEDLINE=95386590; PubMed=7657714;
RA Pal C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;
RT "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar
RT protein p130 are associated with nucleologenesis.";
RL J. Cell Sci. 108:1911-1920(1995).
RN [2]
RP SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).
RC TISSUE=Bone marrow;
RX MEDLINE=96053387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (K1A0001-K1A0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid

```

```

RT cell line KG-1.";
RL DNA Res. 1:27-35(1994).
RL [3]
RP ALTERNATIVE SPLICING.
RX MEDLINE=96205319; PubMed=8630004;
RA Pal C.-Y., Yeh N.-H.;
RT "Cell proliferation-dependent expression of two isoforms of the
RT nucleolar phosphoprotein p130.";
RL Biochem. Biophys. Res. Commun. 221:561-567(1996).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=97168979; PubMed=9016786;
RA Chen H.-K., Yeh N.-H.;
RT "The nucleolar phosphoprotein p130 is a GTPase/ATPase with intrinsic
RT property to form large complexes triggered by F- and Mg2+.";
RL Biochem. Biophys. Res. Commun. 230:370-375(1997).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=20036810; PubMed=10567578;
RA Chen H.-K., Pal C.-Y., Huang J.-Y., Yeh N.-H.;
RT "Human Nop140, which interacts with RNA polymerase I, implications
RT for RNA gene transcription and nucleolar structural organization.";
RL Mol. Cell. Biol. 19:8536-8546(1999).
CC -!- FUNCTION: RELATED TO NUCLEOGENESIS. MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC -!- SUBUNIT: INTERACTS WITH RNA POLYMERASE I.
CC -!- AND WITH CASEIN KINASE-II.
CC -!- CYTOPLASMIC LOCATION: SHUTTLES BETWEEN THE NUCLEOLUS AND THE
CC PRE-NUCLEOLAR BODIES WHICH ARE SUBSEQUENTLY RELOCATED TO NUCLEOLI
CC AT THE EARLY G1-PHASE.
CC -!- ALTERNATIVE PRODUCTS:
CC Name=Alpha;
CC Name=Alternative splicing; Named isoforms=2;
CC IsoId=Q14978-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q14978-2; Sequence=VSP_004338;
CC -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK2 AND PKC SITES. THERE IS EVIDENCE
CC SUGGESTING THAT CDK2 KINASE PHOSPHORYLATES P130 AT THE M-PHASE.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z34289; CA84063.1; -
CC EMBL; D21262; BA804803.1; -
CC PIR; I38073; I38073
CC Genew; HGNC:15608; NOLC1.
CC
CC GK; Q14978; -
CC MIM; 602394; -
CC DR GO; GO:0005737; C:cytoplasm; TAS.
CC DR GO; GO:0005730; C:nucleolus; TAS.
CC DR GO; GO:0007049; P:cell cycle; TAS.
CC DR GO; GO:0007067; P:mitosis; TAS.
CC DR GO; GO:0006364; P:rRNA processing; TAS.
CC DR InterPro; IPR006594; Lish.
CC DR Pfam; PF05022; SRP40_C; 1.
CC DR SMART; SM00667; Lish; 1.
CC DR PROSITE; PS50896; Lish; 1.
CC
CC Nuclear protein; phosphorylation; Repeat; GTP-binding; ATP-binding;
CC Alternative splicing.
CC
CC DOMAIN 10 42 LISH.
CC FT DOMAIN 84 566 11 X 12 AA APPROXIMATE REPEATS OF AN
CC ACIDIC SERINE CLUSTER.

```

```

FT REPEAT      84      95      ACIDIC SERINE CLUSTER 1.
FT REPEAT      125     136      ACIDIC SERINE CLUSTER 2.
FT REPEAT      167     178      ACIDIC SERINE CLUSTER 3.
FT REPEAT      221     232      ACIDIC SERINE CLUSTER 4.
FT REPEAT      264     275      ACIDIC SERINE CLUSTER 5.
FT REPEAT      325     336      ACIDIC SERINE CLUSTER 6.
FT REPEAT      363     375      ACIDIC SERINE CLUSTER 7.
FT REPEAT      425     436      ACIDIC SERINE CLUSTER 8.
FT REPEAT      470     481      ACIDIC SERINE CLUSTER 9.
FT REPEAT      519     529      ACIDIC SERINE CLUSTER 10.
FT REPEAT      555     566      ACIDIC SERINE CLUSTER 11.
FT DOMAIN      68       82      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN      204     382      INTERACTS WITH RPA194.
FT DOMAIN      384     587      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN      601     617      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES     563     563      PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT VARSPIC     241     241      K -> KWTITVSRAE (in Isoform Beta).
FT             241     241      /FTid=VSP_004338.
FT CONFLICT     3       3       D -> A (IN REF. 2).
FT CONFLICT    133     133      R -> S (IN REF. 2).
FT CONFLICT    291     292      YA -> SV (IN REF. 2).
FT CONFLICT    456     456      S -> P (IN REF. 2).
SQ SEQUENCE     699 AA; 73720 MM; DED4AD94EDF659FB CRC64;

```

Query Match 2.9%; Score 206.5; DB 1; Length 699;
 Best Local Similarity 22.3%; Pred. No. 0.014;
 Matches 93; Conservative 55; Mismatches 172; Indels 97; Gaps 13;

```

QY 1045 ENIKOTDAOGPPDARKNEKLYVCYAMNIMSKSTTYSLESPKDYLPARFTQPDKN 1104
DB 228 DDSEEEKAAATPKKTVKPKQVAKAPVKAATPTPKSSSSSEDSSEDEEOK--KPMKN 284
QY 1105 FSNTKNTLPEMKSFPTPGKPKTTNVLGAANKPLSSAGKQSQTKSRMETVSNASSSN- 1163
DB 285 KPGYSYAP-----PSAPPPKKSIG-TPPKKAYEKQOVES--EDSDSDSDSSE 334
QY 1164 -----PSPGRIKGRLDSEMDHSEMDYTNSSPLPGKSKDPKRDSD 1205
DB 335 EEKKPTKAVVSKATTKPPAKKAESSSDSDSSE-DDEAPSKPAGTTKNSSNRPVAV 393
QY 1206 LVRSLELEKPPGRKKTTPY-----TQEEKLGMDLT-KLYQEQKPKGSORS 1249
DB 394 TTKSPAVKPAAPKQPVGGOKLITRRADSSSSSEESSSEEEKTKRMVATTTKPKATAKA 453
QY 1250 RRGHTASESDEQWPEEK---LKEDILENEDENONSPKKGKGRPPKPLGGGP----- 1302
DB 454 -----ALSLPAKQAPGSGRSDSSSDSSSEEEEEKTSKSAVKKKPKQVAGGAPSKPA 507
QY 1303 -----KEEPTWTKSKGSKKKSGPPA-----PEEEE 1328
DB 508 SAKGKAESSNSSSDSDSSEEEEEKLKGKSPRPQAPKANGTSALTAQNGKAANKSEEEE 567
QY 1329 EEEKQSGNTQKSKSKHRVSRRAQRAESPSSAIE-STQSTPQKGRGRPKTPSP 1384
DB 568 EEKKKAHVVSKSGSLKRNQNEAKAEETPQAKKIKLQTPNTPPKRRKKGEEKRASSP 624

```

Search completed: September 24, 2003, 20:17:48
 Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 20:01:29 : Search time 31 Seconds

(without alignments)
4315.182 Million cell updates/sec

Title: US-09-512-581b-2

Perfect score: 7193

Sequence: 1 MAHSKTRTNDGKITPPGVK.....QKGRGRSPSKRPPSPKKNV 1391

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR.76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2698.5	37.5	851	2 T00374	hypothetical prote
2	807	11.2	1579	2 T23142	hypothetical prote
3	632.5	8.8	1506	2 S52957	blind protein - Eme
4	571.5	7.9	1303	2 E96805	hypothetical prote
5	557	7.7	1277	2 S54451	hypothetical prote
6	309.5	4.3	390	2 T43647	hypothetical prote
7	303	4.2	780	2 E96840	hypothetical prote
8	264	3.7	990	2 H86293	protein T24D18.4
9	235.5	3.3	1875	2 S38173	myosin-like protei
10	234.5	3.3	852	2 T06310	hypothetical prote
11	231.5	3.2	2253	2 T30336	nuclear/mitotic ap
12	229	3.2	2663	1 S28261	centromere protein
13	227	3.2	1916	2 S46157	RFL1 protein - yea
14	223.5	3.1	4717	2 T41581	hypothetical colle
15	223	3.1	2139	2 T18296	myosin heavy chain
16	222.5	3.1	1819	2 A71928	cag island protein
17	222.5	3.1	2301	2 T02323	nodulin-like protei
18	221.5	3.1	1927	2 G64585	cag pathogenicity
19	220.5	3.1	1727	2 T50073	myosin-like coiled
20	217.5	3.0	1020	1 OPHU8	neurofilament trip
21	216.5	3.0	1790	2 S67353	transport protein
22	210	2.9	1072	1 A37221	neurofilament trip
23	209.5	2.9	1365	2 T45031	hypothetical prote
24	209	2.9	1242	2 G88480	protein C16A3.7
25	207.5	2.9	854	2 S02003	neurofilament trip
26	207.5	2.9	3488	2 T34418	hypothetical prote
27	206.5	2.9	699	2 T38073	nucleolar phosphop
28	206	2.9	6642	2 T29757	protein UNC-89 - C
29	205	2.8	2094	2 S33124	tptr protein - huma

30	203.5	2.8	2364	2 A56577	microtubule-associ
31	202	2.8	2401	2 T28676	riophy protein -
32	202	2.8	2748	2 S57976	nuclear migration
33	202	2.8	6713	2 B89921	hypothetical prote
34	201.5	2.8	3225	2 I52300	giantin - human
35	200.5	2.8	2269	2 T28677	riophy protein -
36	200	2.8	1837	2 T41023	probable nuclear p
37	199	2.8	1359	2 T34036	hypothetical prote
38	199	2.8	2464	1 QRMSP1	microtubule-associ
39	199	2.8	3187	2 JCS837	364k Golgi complex
40	198.5	2.8	1695	2 T19823	hypothetical prote
41	197.5	2.7	1538	2 T29095	cardiac muscle fac
42	197.5	2.7	3259	1 A56539	giantin - human
43	195.5	2.7	1679	2 S48385	hypothetical prote
44	195	2.7	488	2 T46014	cylicin II - bovin
45	194	2.7	374	2 C88734	protein F32E10.6

ALIGNMENTS

RESULT 1					
T00374					
hypothetical protein KIAA0648 - human (fragment)					
C:Species: Homo sapiens (man)					
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000					
C:Accession: T00374					
R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998					
A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl					
A:Reference number: Z14142; MUID:98403880; PMID:9734811					
A:Accession: T00374					
A:Status: preliminary; translated from GR/EMBL/DBJ					
A:Molecule type: mRNA					
A:Residues: 1-851 <ISH>					
A:Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PTD:g3327110					
A:Experimental source: brain					
C:Genetics:					
A:Note: KIAA0648					
Query Match					
Best Local Similarity 62.1%; Pred. No. 1.1e-121;					
Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;					
QY	477	LYIYATLDLNAVALNEMKCONLHROKDLIDLKOKPTDASVAIRSKVWVITRNL	536		
DB	1	LYIYASLDLPAVALNEMKCONLHROKDLIDLKOKPTDASVAIRSKVWVITRNL	60		
QY	537	PDPGKADPMKKFTQVLEDDKIRKOLEVLVSPSCGKOAEGCVREITTKKLGKPKPTNP	596		
DB	61	PDPGKADPMKKFTQVLEDDKIRKOLEVLVSPSCGKOAEGCVREITTKKLGKPKPTNP	120		
QY	597	FLEMIKFLERIAVHIDTESISALIKOVNKSIDGTADDEDEGVPTDOAIRAGLELTKVL	656		
DB	121	FLEMKFLERIAVHIDTESISALIKOVNKSIDGTADDEDEGVPTDOAIRAGLELTKVL	180		
QY	657	SFTHPISFHSAREFESLACIKMDEKVAEALDIFKNTGSKIEDEPPHRSALLPVLLH	716		
DB	181	SFTHPISFHSAREFESLACIKMDEKVAEALDIFKNTGSKIEDEPPHRSALLPVLLH	240		
QY	717	KSKKGPPOAKVAHCHAFIFSSKEPOFAOIFELHSLDPSNLEHITPLVLTGHTALL	776		
DB	241	KAKRGTHQAKOAVHCHAFIFSKKEPOFAOIFELHSLDPSNLEHITPLVLTGHTALL	300		
QY	777	APDQFAAPKSWATFIVKDLNMDRLGKTKLWVDEVSPEYVVKIOAIKMWVRL	836		
DB	301	APDQFAAPKSWATFIVKDLNMDRLGKTKLWVDEVSPEYVVKIOAIKMWVRL	360		
QY	837	LGMKNHKSQGTSLRLTLITLHSDGLTDEGKISKDMRSLRLAAGSAIVKLAOEPCYH	896		
DB	361	LGMKNHKSQGTSLRLTLITLHSDGLTDEGKISKDMRSLRLAAGSAIVKLAOEPCYH	420		
QY	897	EITLLEQYCALAINDEQYQVQVFAQKLUHKLGRSLRLPLFEMALICALAKDPVIERRA	956		

Dh 1347 GSSAVGKMKRKRKSTKTGDDWSSDGEAGAVASSSSTRRRNRGS-ASRRIYADDDSD 1405
Qy 1261 E--QOMPEEKRLKEDILLENEDQNSPPKKGKGRPPPLGGTPEEFTMTKSGSKK 1318
Dh 1406 EDDMEMELNQRARDEDEGEDQ-----ANDIENGSDLSE-----LSESDSMTL 1448
Qy 1319 SGP-----PAPEEEEEEROSGNTQOKSKS-----KOHRSRRAOQRAESP 1359
Dh 1449 EEPEDDDQPSKEEDQPDQKONGDDAOPASPVASAKAVPGKAMKATLP 1499

RESULT 4

E96805

hypothetical protein T5M16.19 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: E96805

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:1130712

A:Accession: E96805

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1303 <STO>

A:Cross-references: GB:AE005173; NID:96382504; PIDN:AF07790.1; GSPDB:GN00141

C:Gene: T5M16.19

A:Gene: T5M16.19

A:Map position: 1

Query Match 7.9%; Score 571.5; DB 2; Length 1303;

Best Local Similarity 20.6%; Pred. No. 1.2e-19;

Matches 294; Conservative 270; Mismatches 566; Indels 297; Gaps 53;

Qy 38 VVKTFMMDDSEEEKELYLN-----LALHLASDFELKHPGKDYRLVAVCLADIER 89
Dh 21 VANTLSKIDOPSATNKKEGKLIAELRPLKSIKHAHLNKRNDNSLTVVCSELEFR 80
Qy 90 TVAEADYTSPPDKLIDFMETTRQLKGLIEDTSPQFNRYFLENTAMVKSYNICFEL-- 147
Dh 81 ILAPHLPE-EDEYLRDIFTLFIAEFSELDTVSPYFSRAKILETVSLK---FLMLMD 136
Qy 148 EDSNEITQLKRTLFSTYINNGH-----NQR-----VHMHMDLMSII 185
Dh 137 EDCODLVHEMFNMFFSLVREHHOOLINOKSMKTQORRANTQOTHSLENNILAIMSDVL 196
Qy 186 CEGDTVSOQLDLYLVNPAKKNLKNQAYDLAKALKRTPAQALRYPTTFEYFNQVLMCK 245
Dh 197 EE--EANSRYVAVILELVKAGEPTTSGADKLASSLIERCDRIEPLICSLTJSCFMEKD 254
Qy 246 TTSIDSEHVFDELLELYNDSHLLSVLPQLEFKLSNDNEERLQVAVKLAKKFGAKDS 305
Dh 255 SIGTINKDSYHEIIFKISLAPQMLAVIKRLQELLTDQDVAKIKALNLAGRIFAQPKH 314
Qy 306 ELAS---QNKPLMOCYIGRNDIHVPIRLCEVGFASHCLMNH--DIAKDLTEYLKVRSH 360
Dh 315 CLSSYVETQDLYAEFLRRFSKSAEVRMAALCKGCKCYFANPSGNKASGVLAIOERLL 374
Qy 361 DPEAIIHDYIVSYVTAAKDILLVNDHLNFVREPLDRKWRKREKAMGLAIOYKYA 420
Dh 375 DFDQVRYQALVACDITKKNMKYVPLNLISSEASERDKRIKSYRKALOKLTVEYDYC 434
Qy 421 LQSAAG---KDAKQIAWIKDKLHLHYQNSID--RLVERIFAQYVMVHNLETERM 474
Dh 435 DKQSEGMITTDNEQIPLC---KILLICCEKNCEFRSQNLLELYSLDLPFRLLPVERM 491

Qy 475 KCLYLYVATLIDLVNKALENMWMKQNLRHQVKDLIDLKQPKTD-----ASVKAIFSK 528
Dh 492 RHVVOCFAMNHIIHLKSLNSILSQRRRLQNLRLHOLMRKAVDNIDEAQKKKSYEVK 551
Qy 529 VAVITRNLPDPGKQDFKFKFTQVLEDEKIRKQLEVLVSTYSCQKQAGCVREITKLG 588
Dh 552 ---LSACPDASEEDLEFKIDRM--RDAISFDVLTLLLE-ELSTNAQIIKEKFLKMG 605
Qy 589 NPKQPTNLFIMIFLERIARIVHIDPTESIALIKOVNKSIDGADDEDEGVPPDQAIRA 648
Dh 606 ---YKHSLEFELRLSTKCSPTSFSSHHVCLLNQLCGSTSANT-----QLKAP 651
Qy 649 GLEELKVLSTPHIPSFAETFEESLACLKMDKDEVAEALQIFKNTGSKIEDEPHTRS 708
Dh 652 SIKLLVLINMFPSYLRGSE--KQFLKLE-ENSDADELIVLSKAPYSVNF---G 704
Qy 709 ALLEVLNHSKSKGPPRQAKVAIHICHAIF-SKETQFQRIEPLHKS-LDSNLEHLTP 766
Dh 705 DYPLVLEKVCLEGRSQTKCAVSAISLAGSSEKSVSELCMLMSILCGRNIPPTLOS 764
Qy 767 LVITGHIALAPDQFAAPKSWVATFIYKDLMDRLPGKRTKLWVDEEVSPEYTWYK 826
Dh 765 LACVQYSVLEYDN-----IYEDI-----TSYIY-----RVYQI 793
Qy 827 QAIKMYVRLWMKNNHNSKSGSTPLRLITTLIHSDDG-----LTEQKISKDPMRRL 879
Dh 794 YGLTLYVKSFLPRHQVYVRKIDDLINLIKTKLSQGHGKISMCLEFLVEICSEPTGANVR 853
Qy 880 LAAGSAIVKIAQEOCYHEIITLEQYQALAINDECYOVORVFNOKLH--KGLRLRLPL 937
Dh 854 LAAKAAVILLSRKDLH-----ISPEVRLTILMAKSFYINGF----- 892
Qy 938 EYMAICALCANDPYKERRAHAROCIVKNINVRREYLKQHAASSEKLISLPEYVVPYTIH 997
Dh 893 -----INKATRESRTC-----BDLDGSESLTDS-----PYMTVFLIH 925
Qy 998 LLADHPDYVKQDIEQLK---VKEC--LMFVLEILMAKEN-----NSHAFIRKAYEN 1046
Dh 926 VLADHPPEPS---EDCRDEHIVARFCGPFLPSVLQVLLSINNFTTIKETAPFLCIPRA 981
Qy 1047 IKOTKDA---QGGDDAK--MNEKLYVCVAAINIIMSKSTYSLESPPKQVLPARFT 1099
Dh 982 IKRAEDAVDSKRTNKAKSRIRNALBOSFETIRYIHROSQISMDOQCQKSLA----- 1035
Qy 1100 QPDRNFSNTKNVLPPEMKSFFTPEKPTTVNLAGVANKPLSSAGROKTSRMEF----- 1154
Dh 1036 ---VGSDEKVLPLL-----GNQIETSTG-----STEAQNNTRCSRKRTHLGEH 1078
Qy 1155 -----VSNASSSNSSPGRKIGRLDSSCMD 1180
Dh 1079 ISCNLSLRIVESEIPIKLERHTTCAKESYKASVSNKITSKHS--GVVASLKDIS--N 1134
Qy 1181 HSE---NEDYTMSSPLDG-----KKSQRDSDLV---RSELEKPRGRKKTVTYDDE 1227
Dh 1135 HGEAIIIGRIKILSPDCCFPGYVERKNSNSHKITTFDNGDVELVCLDSESEWTSHE 1194
Qy 1228 KLGMDDLTKIVQEOKPKGSQRS--RRGHTTASEDEQOMPEEKRLKEDILENEDQNSPPK 1286
Dh 1195 SMGOGE--RLGKEHESYSGRNCVPEISHTLAKVNAOK--QTTTKQKKKVPKALNPPAA 1250
Qy 1287 KGRGRPPKPLGGSTPKE-----EFTMTSK-----GSKRKS 1319
Dh 1251 KSKKGNDS--GEGSVSEVTDTSNIKSECLRIKKSIAQFGTKRKS 1295

RESULT 5

S54451

hypothetical protein YMR076c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YM9582.01c; hypothetical protein YM9916.15c

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999

C:Accession: S54451; S52836

R:Gentles, S.; Bowman, S.

submitted to the EMBL Data Library, May 1995
A:Reference number: S54451
A:Accession: S54451
A:Molecule type: DNA
A:Residues: 1-682 <GEN>
A:Cross-references: EMBL:Z49259; NID:9807956; PIDN:CAA89222.1; PID:9807957; MIPS:YMR076C
A:Experimental source: strain AB972
R:Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, April 1995
A:Reference number: S52814
A:Accession: S52814
A:Molecule type: DNA
A:Residues: 659-1277 <PEA>
A:Cross-references: EMBL:Z48952; NID:9763008; PIDN:CAA88801.1; PID:9763023; MIPS:YMR076C
A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:PD55
A:Cross-references: SGD:S0004681; MIPS:YMR076c
A:Map position: 13R

Query Match 7.7%; Score 557; DB 2; Length 1277;
Best Local Similarity 20.3%; Pred. No. 5.8e-19;
Matches 272; Conservative 267; Mismatches 611; Indels 190; Gaps 46;

QY 12 KITPPGKKEISDK-ISKEENVRLKVVKTFMDMDODSEEEKEL-VLNLALHLASDFE 68
DB 8 KLRKNSPIISTSDOLISTNEILDRKALHEELASLDODNTDLTGDKTRDA---LVSKRL 64
QY 69 LKHGKDVRLVACCLADIFRIYAPEARVSPDKIKDIFMTROLKLEDTKSPQFNRY 128
DB 65 LKHGVDGIRAFPTACCLSDILRLYAPADAYTDA-QLTDIFKLVLSQLGQDQENGHIQ 123
QY 129 FYLLENIAMVASYNICELEDSNELFTQLYR-----TLFSVYNNGNQVNHMM 177
DB 124 TYLLTKLEYRSIVLLADLPSSNNLLELFHIFYPDNKSPARLFNVIG----- 173
QY 178 VDLASIIICGDFVSOELDTVL-----VNLVPAHKNLNGQ-AYDAKALKRTQAQI 229
DB 174 --ILGEVISEDSVPLEVRLIFPKFLYNNNEIEGALNVSDCYEVSLLICDTYSNRM 231
QY 230 EPTVTTFFNOVLM-----GKTSISDSEHVFDLILELYNIDSLLSVLPQLEFKLSN 284
DB 232 SRHLTKYSEIIEHATNDNNSRLTVVVKLHLVRLMETVPELINAVIGFIYHELSSE 291
QY 285 DNEERLOVYKLAMFPGA-KSELASQNKPLMOCLGTFNFIHVRLECYKFAHCLMN 343
DB 292 NELFRKEATKLIIGILTSYDNLNFVSTHDFKAWISKIADISPDVREWTESTIPQIIAT 351
QY 344 HPDLAKDLTEYLKVRSHDPEAIRHDVYSI---VTAKKDIL-LVNDHLNLFVRERT 397
DB 352 REDISKELNQLATFIDSDPRVVRTSMIFNKVPVTEIWNKITKAIYTSILHARE-- 409
QY 398 LDKRWVRKEMAMGLAQIYKKYALQ---SAAKDAQAKIAMIKDLHIYYQNSIDRLL 454
DB 410 --KKEVELEICINMAKFSNSLNEIERTYONKEIMEIIDIIPSTLYMYIINDLINEQ 467
QY 455 VERIFAQVMPHNLETTERRMKCYLYVATLIDLVKALNEMWKCQNLRLHQYKDLIDLK 514
DB 468 VDSVTFETLDFEPDNDKRVHRLTLVSHDFDKAFTSPFANARQIKISPAISKYIDSK 527
QY 515 QPKTDAVKAIFSKVMV-----ITRNLPDPGKADPFMKKFVQVEDEKIRKOLEV 565
DB 528 FLNNGEASSSQGRIYNNKYNQTLQWLASGLSDSTKALDALETIKQF--NBERITYLNA 585
QY 566 LVSPFCSCQAEGCVREITTKLGNPK-----QPTNPLEMIKELLERIAVH 612
DB 586 CVFNIDIFLTFKNCYNELVSKIQTPGLFKKYNISTGASIMPRD-TAKYIQILIFRASPDI 644
QY 613 IDTISALIKQVKKSIDGTADDEGVPDQAIIRAGLE--LTKVLSSTHPISE-HSMT 669
DB 645 YNVSNISVLLNLSNS-----DAKQDLKRRLLDDISVYNPTLFEQDQIRT 689
QY 670 FESLLACL-----KMDKVAEALAIQIFKNTGS-KIEDEPHIRSLPVLNHRKSKGP 722

DB 690 LKTIKDDDDPAEKENMLSLEALKTYLKASKTLKQVDFFD--TFEFTKLYDPAVESK 747
QY 723 PROAKYAIHCHAIHAISSKETQPAQI---PEPLHKLSDNSNHLITPLVYIGHALLAPD 779
DB 748 PEITFYAAKRLI-ALSPKAEETLKIKIRILPLDLQKDYFTSHIILVMEIFKFPVHLD 806
QY 780 QFAAPKSMVAFTYIKDLMNDRLPGKKTTRYK-WVPDEVSPEYMW----KIOAIKMMVR 834
DB 807 D-----STDIIISYLKEVLLSNQVGDSSKKELDWEDSLSDTKSAIGNKVFYTKLFTN 861
QY 835 WLGM-----KNHNSGCT-STPLRLTYTILSDGL-TEOGKISKPDMS---RLRLAG 883
DB 862 KLRSLAPVPRDLAESFTETKMKLFYLLASGGLISBPKNKEFYPTPSNVQTKLRVAG 921
QY 884 SAIVKLAQPCYHEIITLEQYQCALAINDECYQVQVPAOKLHGLSRHLPLEYMAIC 943
DB 922 IOVELKARISNINNEIKPSDIIKLINLVEDESLPVKFTLEQLKDYVANELISIKFLDLY 981
QY 944 ALCAKDPYKERRAHARQCLVKNINVRREYLKQHAVSEKLSLLEPYVVPYTHLLADP 1003
DB 982 FFTAVERPVELKTYTKIMI--NFTGLKSPFK-----GTIERALPRLIHAIHNP 1030
QY 1004 DYVKQDIE-----QLKDVKECLFWLEIILMAKNENSHAFIRKVENIK-----OTK 1051
DB 1031 DIVGSLDESGDAYLNALTTAIDYLLFYDSTIAQGNFSLIYLSRVKNYQDKLVEDELD 1090
QY 1052 DAQGPDDAK-----MNEKLYTVCDVAMNIMS---KSTYS-----LESPPDVLPA 1095
DB 1091 EEEGQKEAEARKHNRPYQCKMYIIGELSOIMILNKEKKNNQMSAYPGKLMIPSLFKP- 1149
QY 1096 RFTQPRDNFNTKYLRLPREKSPFTPKPKRTYTNVLGAVNKPPLSAGKOSQSKSRMEYV 1155
DB 1150 -FATVQEAQLS-FKYTIPESL-----EKIONNKAKIGRLHSTQROQLKRLAL 1201
QY 1156 SNASSSNPSSPGRIKGRIDSSEMDHSENEYTMSPLPGKKS-D-KRDSDLVRSLEKP 1214
DB 1202 ENNESQKKKKKYNHARQADDEGDDGDESDDSDSYSPSNKNNEFKKGHENY--MKRL 1258
QY 1215 RGRKTPVTEDEEKLGMDDL 1234
DB 1259 RVKKEVDYKDDDED--DDI 1274

RESULT 6
743647
Hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T43647
R:Lee, M.; Yoo, H.S.; Chung, K.S.
submitted to the EMBL Data Library, February 1998
A:Description: Clone 17 (blnd homologue).
A:Reference number: Z22596
A:Accession: T43647
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-390 <LE>
A:Cross-references: EMBL:AF049529; PIDN:AAD02493.1

Query Match 4.3%; Score 309.5; DB 2; Length 390;
Best Local Similarity 24.8%; Pred. No. 8.4e-08;
Matches 105; Conservative 76; Mismatches 192; Indels 51; Gaps 11;

QY 825 KIOAIKMMVRMLGKKNHS--KSGTSLRLTYTILSDGL-TEOGKISKPDMSRLDA 882
DB 4 KVLAIKRVLYNRLRAAGSTALNIGAPIIKLKYLMADGETSPKNTKISRATRLTA 63
QY 883 GSAIVKLAQPCYHEIITLEQYQCALAINDECYQVQVPAOKLHGLSRHLPLEYMAI 942
DB 64 SKYFLKLCISIPYAAHIDPSSVSNLLTMPDENFVRNMLFKTLQKQLOKLLKPLSYTP 123
QY 943 CALCAKDPYKERRAHARQCLVKNINVRREYLKQHAVSEKLSLLEPYVVPYTHLLAND 1002

```

Db      124 LFLTADPEEETITKA-----SIRKSOVAFFOKTHDFTMYVATYTLHLHSH 172
QY      1003 PDYVKYQDLEQLDKVECLMFVLEILMAKNENNSHAFIRKKNVENIKOTKAOQDPDAKM 1062
Db      173 PDLSTSESESLDFIAIRFYVDTYVNSEV---PIVFLMQRIKSYDV--IEDG--N 224
QY      1063 EKLYTCDAVAMNIMSKS-----TTY--SLESPKDPVLPAREFTQDPDKNFSTKLYLP 1113
Db      225 NRYIVLSDMAOKILOVKSQNFQMSLTJTYPOIKILPILRPIPSIDKKRIFN--KIFIT 262
QY      1114 PEKSFPTPGKPTTNVLGAVNKPLSAGKOSOTKSRMETVSNASSSSPSPGRIKGR 1173
Db      283 PKHES-----QIEHAIKTPVSEFAKOTTNKHANLKOCTHSSKSDKSSRRK-- 330
QY      1174 LDSSEMDHSENDYTMSSPLPGKSKSDRDSDLVRSELEKPRGKKTPVTEQEEKLMD 1233
Db      331 ---NEKRRKLNQNPNNRNVPERSSSRFQGIKIRINYSAPS-----SSEIESEEEIESED 383
QY      1234 LTKL 1237
Db      384 FDEI 387

```

RESULT 7

F96840
 hypothetical protein F23A5.16 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 A:Accession: F96840
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huiztar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: F96840
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-780 <STO>
 A:Cross-references: GB:AE005173; NID:96503292; PIDN:AAF14668.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F23A5.16
 A:Map position: 1

```

Query Match      4.2%; Score 303; DB 2; Length 780;
Best Local Similarity 20.4%; Pred. No. 4.3e-07;
Matches 153; Conservative 118; Mismatches 254; Indels 224; Gaps 27;

QY      18 GYKEIDSKISEEMVRRLKAVVTFPMDMODSEEEKLYNLAL-----HLASDFLKH 72
Db      18 GTNLSPSPSTDDLTLTLDFTESLKNVED---OPLSMQSAALIPSRNALVSDLLSH 73
QY      73 GKDVRLVACCLADIFRIYAPAPYTPDKIDIFMFTITROLGLEDTKSGPOFNRYFL 132
Db      74 DSDVRSVSVSCLFEIRITAPERPYSDDLAKETFLITIAFEKLADASSRKYKAFVL 132
QY      133 ENIAWVKNYICFELEDSNEIFQLYRTFESVINGNHNOKVHMHWDLMSIIEGDTVS 192
Db      133 DINAVKKSCILVMDLE-CYDILLQMFNRFKFIKSDHPQLVFSMELIMAIIDETPOVS 191
QY      193 QEILDVYVNLVPAHKNLKNQAVDLAKALKRTAQALEPTITFFNOVYMLGKTSISDS 252
Db      192 TDLDSLALTFVKKNQNVSPMSWSLAEKVLSCARKKPYIT---EALSRGTSIDMS 247
QY      253 EHFEDLILLEYINIDSHLLSVLPQLEFKLSNDNEERLOY-----VKLAK----- 298
Db      248 PVVSSICQSVFN-----TPKVISPVNTRKEHESLKYQVRSESTDAETIGKGRKPN 298

```

```

QY      299 -----MFGAKDSELASQNKPLMQ-----CYLGRNDIHVPIRLCYKFAHCLM 342
Db      299 MNPEVDYISMLSGKRDPLKTSNNKKIQKSGGVSLGKVPAKTLPPE----- 348
QY      343 NHP-DLAKDLTEYK-----VASHPEE----- 364
Db      349 NSPATSSRLTGLSKRSRVKMDSEYDSDLSLSPRLKLASCDFDEPNEDDRKIGNSS 408
QY      365 ---AIRDVIYSIYTAARKDIL--LVND-----HLNPFERITDRKWRVREKAMGLA 413
Db      409 KQTRSKGLEKSOQTAKKVEKVEAKIYNSSGKRLSANSVAKRRLE---RAPDITLPOS 465
QY      414 QIYKRYALQSA-----GKDAKO---IATIKD 439
Db      466 SKRKWVSQVAARLANESEEPFKSHPTRRRYRKFEVSDGFCEDLVGRVNIWPLDK 525
QY      440 LHIYYONSIDDLVLRITAOYVWPNINLETTEMKCLLYLYATLDINAVKALNEMKQ 499
Db      526 ---TFEGVID-----SYCTKKMHRVYISDGSDEELNTEERN-- 561
QY      500 NLIRHGVKDLIDIKQKTDASVKAIRSVWITRNLPDQKADPFKKFTQVLEDEKI 559
Db      562 -----ELLEDTSADEKEL-----DLPESTPLSDIMQ-----QKV 593
QY      560 KQOLEVLS--PTCS-----CKQAGCVREITKRLGNPKOPTNPFLEMIKFLERIA 609
Db      594 KSKNVAIVSEPTSSGVRSSRTLMKDKGKRLNKQVETKREKN-----LRSLK 644
QY      610 PVHDTESISALIKQVAKSIDGTADDE 638
Db      645 ELNAETDRTA--EEQEVSLAESDDRSE 670

```

RESULT 8

H86293
 protein T24D18.4 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 A:Accession: H86293
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huiztar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: H86293
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-990 <STO>
 A:Cross-references: GB:AE005172; NID:96587800; PIDN:AAF18491.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T24D18.4
 A:Map position: 1

```

Query Match      3.7%; Score 264; DB 2; Length 990;
Best Local Similarity 17.5%; Pred. No. 4.3e-05;
Matches 241; Conservative 188; Mismatches 440; Indels 506; Gaps 50;

QY      43 MDMDDOSEEEKELYNLALSLASFELKHGKQRYLVACCCLADIFRIYAPAPYSPDK 102
Db      47 VEDQLSSVOKALHPRALVAD-LRNPDSDVRSVSVSCLTEIMKITAPEAPY-NDQ 104
QY      103 LKIDFMEITROLKLEDTKSGPOFNRYFYLLENIAWVKNYICFELEDSNEIFQLYTTF 162
Db      105 MKDIFQVITIAFEKLADASSRKYKAVILEYTAQVASSLVMDLE-CDLVLEMPQREL 163
QY      163 SVINGHNQVHMHWDLMSIIEGDTVSQELDITVNLVPAHKNLKNQAVDLAKALL 222

```

```

Db      164 KIIRDPHQLVLSHETIMIVIDESEVPDMLLELITLYTKKSDQVSPALTYEVKL 223
Qy      223 KRTAOALEPYITTFNOVLMLGKTSISDLSHVDFDLELYNIDSHLLSVLPOLF- 279
Db      224 SSCCTCKLOPCIM-----EALKSSGTS-----IDMYS-----PYVSSICQSEFATY 263
Qy      280 -----KLSKNDNEERL---OVV---KLLAKMFGAKDSFLASQNKPLMOCYLGREFNDIHP 328
Db      264 QAHNDVPRKONEADEKISEGQVPSDLSLEDKLNLGLSKRGTRKRSASGTRRANGD--- 320
Qy      329 IRLCEVFAASHCLMNHDPDLAKDLELYLVKRSHPDEEARHDIYVISAIPAACKDILLVNDH 388
Db      321 ---EKVITANGCLSESDT-AETASGSTKRGKMKPRSLMNPREGVSFKTSSSKV----- 370
Qy      389 LINFVREBTLDKRMVRKREAMMGLAQYIKYALOSASGAKDAKQIAMIKDLHLHIYONS 448
Db      371 -----OEKELG-----DSSLCKVAACKVP----- 389
Qy      449 IDDLVLVERIFAQYMWPNHLETERMKCLYLYLATIDLNAYKALNEMMKQNLRHQYKD 508
Db      390 -----LPSKVQOTNO-----SVVISLSSSGRAFTGSKRSR- 420
Qy      509 LLDLIKOPKTDASVKAJFSKVMYITRNLPDGAQDFKFTQVLEDEK--IRKQLEVL 566
Db      421 ---TKMEETHDVSSVATQ-----PAKKOT-VKTNPAKEDLTKSNVKKHEDGI 465
Qy      567 VSPFSCCKQA--ECGVREITKLGKPNKOPTPPLIE--MIKLLERIAPHVIDTESISALI 622
Db      466 KTGSKRSKEKADNLAKTSARK-----PLAETMVMVKPSGKK--VHSDAK----- 508
Qy      623 KOVKSIDGTADDEGVPTDOAIRAGLELKLKVLSTFHPISFHSAEFTESLACIKMDE 682
Db      509 ---KKNNEGASMD---TRIPQSSKS-----KKKDS 532
Qy      683 KVAEALQIENKNTSGKIEEDPHIRSALEPVLHHSKKKPPROAKYAIHCIAHJFSKET 742
Db      533 RATTPARK-----KSEQAPKSHPKMR-----IAGEV 560
Qy      743 QFAQIFERLHSLDPSNLEHLITPLVYIGHIALALAPDOFAFPKSWATFYKDLMDR 802
Db      561 E-----STNTEL-----GEE 570
Qy      803 LPGKRTTKLVNPDDEVPETWVKIOATKMWVRWLLGMKNHNSKSTSTRLTLTLHSDG 862
Db      571 LVGRVAVWMPDLKKFEGVAKSKCYKAKM-----HQ-----VTYSDG 608
Qy      863 DLTEOGKISKPDMSRLLAAGSAIVKLAQEPYHEITLLEQYOLCALAINDECYOVROVF 922
Db      609 DV-BELMLKK--ERFKLIEDKSSASEDKEDDLLESPLSAF----- 646
Qy      923 AQKHLKGLSLRLLEIYMAICALCAKDPVKERRAHARQCLKYNINVRREYIKOHANVSEK 982
Db      647 -----IOREKSKKRIYK 661
Qy      983 LLSLLEPVVYVYTIHLAHDPDYKVDIEOLKDYKCECLMFLVLEILAKKNNSHAFIRK 1042
Db      662 ---VER-----SSPEVSSMOQTMKKD-----S 682
Qy      1043 MVENIKOTKQAQGPDAKMKNEKLYTCDVANNIITSKSTYSLESPPKDPVLPAFFTOPD 1102
Db      683 VTDSIKOTKRTKGLAKLVNNEP-----ESTTGKLNLSLK-----KLNGBPD 723
Qy      1103 KNFSNTKNLYLPEPKSFTTPCKPKTTYNVLGAVNKPPLSAGQSQTKSSRMETVSNASSS 1162
Db      724 KTRGRT-----GKKO-----KVTOAMHRIK 745
Qy      1163 NPSPGRIRGRLDSSMDHSENYDTMSSPLPGKSKDPKRDSDLVRSLEKPRGRKKTVP 1222
Db      746 -----DQDEDELETKDEBDSKL-GKESDA--EPDMEDHQLPEPHNNVETK 790
Qy      1223 TDEEKLGMDDLTKLVQEQPKGSQSRKRGHTASSESDQWPEPKRLKEDILE--NEDE 1280

```

```

Db      791 TDGEE-----QEAKEPPTAESKTNCEEPNAPPETDQGEKHSKLEKNAPKSDGE 839
Qy      1281 QNSPPK-----GKRGPRPKPLGGTPEKEPTMTKS-----KKGSKKSGPPAPEEE 1328
Db      840 EOEAKEPNAELKTDGENOEAKELTAIRKTDDEEHKVADEVOKSKETVVEPEAGEE 899
Qy      1329 EEBRQSGWTEQSKSKOHRVRRRAQOAREESSASIESTQSTPOKGRGRPKSTS 1383
Db      900 QKSVPEPNAEPKTK-----VEEKESAKQOTADFKLIEKEDMSKTGEEDIKETYS 949

RESULT 9
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR095w; protein YKR415
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: EMBL:Z28320; NID:q486586; PID:q486587; MIPS:YKR095w
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Teast, G.; 1349-1354, 1993
A>Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:q450550; PIDN:CA51948.1; PID:q450554
A:Experimental source: strain S288C
R:Kobayashi, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A>Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549; PMID:8483450
A:Accession: S31207
A:Molecule type: DNA
A:Residues: 1-300, 'A', 302-1875 <KOB>
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C:Genetics:
A:Gene: SCD:MLP1
A:Cross-references: SCD:S0001803; MIPS:YKR095w
A:Map position: 11R

Query Match      3.3%; Score 235.5; DB 2; Length 1875;
Best Local Similarity 19.5%; Pred. No. 0.0023;
Matches 301; Conservative 246; Mismatches 602; Indels 391; Gaps 72:

Qy      19 VKEISDKISKE-----EMVRRLKVVYKTFMDQDQDSEEEKL-VLNLALHLSDFLKH 71
Db      278 IKGLSDSLNSEKQFSAEMSLKORLVLDLESQLANVAVEELNSIRELNTAKYIADSKQY 337
Qy      72 PG-----KDVALL---VACCLADIFRI-----YAPEARYTPDKLQDFMTITQOLKGL 118
Db      338 PENEDLKEQLQTEKLAQCEKECECLRSSITDEADENENLSAASSDFFILKQQLKER 397
Qy      119 DTK---SPOFNRYVYLLFN-TAWKSVYICFELDSNEIFQLRFLFSVNNCHNOKVH 174
Db      398 RTKEHLNQIQETFIYELKHPPIINSFR---ERDMDLENNALNALLLEHNSNEKNKVK 454
Qy      175 MHAVDLMSITCEBD--TVSOELLDPYLVNLVPAHKNLNKOAYDLAKALLKRTQAILEPY 232
Db      455 ELNAKNQILVCENDLQTLTKORD--LCROIQVLLITNSVNSDSKGLRKEEIQ----- 507
Qy      233 ITTFNOYMLMGKSI--SDLSHVDFDLELYNI-----DSHLLSVLPQLEKKSND 285
Db      508 ---FIONMOEDSTITRESQKVVATELVEFKNIIOLEQKNABLLVVRNLADLSKE 564
Qy      286 NEERLQVVKLLAKMFGAKDSFLASQNKPLMOCYLGREFNDIHPVIRLCEVFAASHCLMNH 345

```

```

Db      565 KSKSGSLQKI-----ESETVNEAK-----EAITLTKSEKMLEERI-----E 601
Qy      346 DLAKDELTEYLKSHDPEEARIRHDVIYSITAAKADI-LLVND---HLNFEVERTLDR 401
Db      602 ELQKELEB-LKT-SVNEEDASYSNVITIKOLTEKTRDLESVOYOLQIRISITEST-ENM 658
Qy      402 WVRKREPMNGLAQIY-KKVALQSAAGDAKQI-AMIKDKLHIYYONSIDRLVERIF 459
Db      659 SLINKE-----IIDLQYDSKDISIKLGKESKRILAEERFKL-----SNTLD----- 701
Qy      460 AQYVNHNLLETTERMKCLYLYATL---DLNAVKALNEMKCONLHRQYKDLIDLKOP 516
Db      702 -----LTKANDLRRFDYLONTILKODSKTHEITNEVYCSKSISETLELNT--- 752
Qy      517 KTDASVKAIFSKVWVITRNLPDGGKADQEFMKFTQVLEDEDEKIRKOLEVLPVSTCSCK-- 574
Db      753 KEQOKLRVHLEKRLKQELNKLSPK--DSLRIWTOLOLOKEREDELLSETRKSCQKKID 810
Qy      575 QAEQVREITKLGNP-----KQPTNPLEMIKILLERIAVH-----IDTE 616
Db      811 ELEDALSELKETSOKDHIKOLEEDNNSNIEMWQNKIEALKKDYESVITSVDSKQTDIE 870
Qy      617 SISALIKQVVKSID-----GTADDEDEGVPTDQAIRAGIELLKVLSFTPIPSHAEFT 670
Db      871 KIQYKVKSLKEKTEEDKIRLHTYNAVDET-NDDSLKELKEKSKI-NLTD-AVSOIKY 926
Qy      671 ESILACLKMDKDEKVAEALQIFKNTGSKIEDEPFHRSALLPVLHNHS----- 718
Db      927 KDLY-----ETTSQSLQOITNSKIDSEKFDPTNOKNLTDKETSLEDKISLKEQ 975
Qy      719 -----KQGPRAQ--KYAIHC-----IHAIFSKETOFAOI-----EPLH 752
Db      976 MENLNNELDQKGMKEKADFFKRISILQNNKKEVAVKSSEKSLKIQNDLDOQIT 1035
Qy      753 KSLDPSNLHLITPITIGHIALLRDPQAFAPKSWATP-IYKDLMLNRLGCKTKL 811
Db      1036 ANPANNYBOELQKHAADVSKTISELRQDHT-KYGOVKTLNLSRDQENAL--KENEKS 1091
Qy      812 WVPDEEVSPE-----TWVKTQAIKMMVRLGKNNHNSGSGTSLRLTLTIL 858
Db      1092 WSSQKSLSEQLDLSNRIEDLSSQNKLLYDQIQTAAKKEVNNNSINGGLNITLTLR 1151
Qy      859 HSGGDLTEQGKIKPDMKSLRLAAGSAIYKL--AQEBCHEIITLEQYQALALINDECY 916
Db      1152 RERDIDITKYVAERAKMLROKISLMDVELQDAFTLDNSRVEKENHSSIIOQHD-- 1208
Qy      917 QVQVQVAKRLHKLRLRLPLEYMAJCALCAKDPVERRAHARQCLYKKNINVRREY--- 972
Db      1209 -----IMEKLNQNLN-----LRESNITLKNLENNNNKKKELOSEL 1244
Qy      973 --LKOHAAYSEKLLSLPEYVYPTIHLAHDPRYKVQDIEQLDKKECLMEVLEI--- 1027
Db      1245 DKLKQVAPRLESLELAL-----KYSQKEQELKLAKL-----EVHRW 1282
Qy      1028 -----LMAKNENNSAFTIKM--VENIK---QTDADAGPD-DAKKN-----EKLYT 1067
Db      1283 KKRSDIILEKHEDLSSDEKLESEIENLKEELENEROGAEAEKFNRLRROAQERLKT 1342
Qy      1068 VCDVANNTIMSKTSTYSLSPKDPVLPARFPTQPDKN-PSNTKNYLPRPEKSFPTGPKR 1126
Db      1343 -----SKLSQDSLITQVNSLRDA-----KNVLENSLSEANARIEELQNAKVAQ 1385
Qy      1127 TTNVLGAVVKPLSSACKOSQTKSSRMETVSNASSSSNPSPGRIGKGLDSEMDHSENE 1186
Db      1386 GNNQLEAIRKIQDADAKASRELOAKLE-----ESTSYESTINGL-----NEE 1428
Qy      1187 YTMSSPLPGKSKDKRDSDLVSELEKPRG-RKTPVTEQEELKGLMDLTKLVQEQKPKG 1245
Db      1429 IT-----TLKEIEIKORQIQOOLQATISANEQ---NDLSNIYESMK--- 1465
Qy      1246 SQRSRRKRGHTASDQOWPEEK-RLKEDILLENDEQNSPP-----KKGKRGRRPKR 1296

```

```

Db      1466 -----KSEEDKIKFIKERTQOEYNEKILEAQERLNPNSINNEELKKWSEHDE 1516
Qy      1297 IGGGTPEKEEPTMK-----TSKKGSK-----KSGPPAPEEEEEEER-----QSG----- 1335
Db      1517 VSKIRAEAEALKKRIRLPTEEEKINKIKIEKKKELEEFPEKKEERKSKSMEQSGELDVVL 1576
Qy      1336 -----NTEQSKSKOHRVSRRAQQAES-PESSAIESTQ 1368
Db      1577 RKOLEAKVOKOKELEENYKMKLOEELKQVDPHSSHISDDE 1616

RESULT 10
T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06310
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyssehaert, C.; Dasseville, R.; De Clerck, R.;
  ewes, H.W.; Mayer, K.F.X.; Schueller, C.
  submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15589
A:Accession: T06310
A:Molecule type: DNA
A:Residues: 1852 <BEV>
A:Cross-references: EMBL:AL049607, GSPDB:GN00062, ATSP:F11C18.80
A:Experimental source: cultivar Columbia; BAC clone F11C18
C:Genetics:
A:Gene: ATSP:F11C18.80
A:Map position: 4
A:Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3

Query Match      3.3%; Score 234.5; DB 2; Length 852;
Best Local Similarity 16.9%; Pred. No. 0.00092;
Matches 236; Conservative 175; Mismatches 391; Indels 591; Gaps 49;

Qy      18 GVEISDKISKEMVRRLKAVVVTFFMDMDSEEEKELYLN-LALHIASDFLKHPEKD 76
Db      16 GERLIDPPSSLDLSFLDKLFVSLAEVQSPDSMOMALTPLMKGLVGGKLRKHSDDV 75
Qy      77 RLIVACCLDIPIRIYAEAPYTSPODKLIDFMFTRLQKLEDTKSPQFNRYFYLLENIA 136
Db      76 KVAVAAICISITITAPDAFY-DDDQKKEFKLIVSFEELVDKSSSYAKRISILFTVA 134
Qy      137 WKSYSNICELEDSNEITQLYRTPFSVINNGHOKVHMVMDLSSIIICEGTVSQELL 196
Db      135 KVRSCVVMIDLE-CDALLIMFQHLKAIIRDHSGNVFSSMENTMFLVLESBDISEML 193
Qy      197 DTVLVNLYRAHKNLKNQAYLAKALKRTQAIEPTT-TFFNQVLMLGKTS--ISDLS 253
Db      194 SPIL-HSVKKDDDEISQVSRRLAQVLSNCASKLKTLYLTVAVKSSGVPLDKYSNIVASICE 252
Qy      254 HVPDLLELYNIDSHLLSVLPQLEFKLSNDNEERLQVYKLLAKMEGAKDSELAQNK 313
Db      253 GTF-----SALOQ--DQVANEKED-----SQG--- 273
Qy      314 LMQCYLRENDHIVPRLCEVAFASHCMLNHPDLADLTEYLKVRSHDPEAIRHDVIVS 373
Db      274 -----HIKRETEVEKAAEISTPERTAPRDESGKSGVSN---GVAQOONSVD 318
Qy      374 IYTAARKDILLVNDHLNFPRERTLDRKRVKREAMMGLAQIYKKVALQSAAGDAKQI 433
Db      319 TDSMKKQD-----DTGAKDEPQOL 337
Qy      434 AMIKDKLHIYYONSIDRLVERIFPAQVWVPHNLLETTERMKCLYLYATLDLNAVAKLN 493
Db      338 -----DNPRTNDLNNITBEKDFV-----HQIEKEKNESS---SVKQADLSKSDIK 381
Qy      494 EMKCONLKHQYKDLIDLKQPTDASVAKFSKYWVITRNLPDGGKADQFMKKFTQVYL 553
Db      382 EETPEPELL-----DSKVDITSPVDSVTAANS----- 410
Qy      554 EDEKIRKQLEVLVSPQSCQA-----EGCVREITTKLGKGNKQPTNPLEMIKFL 604

```

Db 411 SENEK-NKSVQILPSTSGDETANVSPSPMAEELPEQSPVKTKTANOKKSS----- 461
QY 605 LERLAP-VHIDTESIALIKOVNKSIDGTADDEGCVPTDOAIRAGLELLKVLSTHPIS 663
Db 462 TEEVPPSPASIAITEEVS----- 477
QY 664 FHSAPETFSILACLMKDEKVAEMALQIEFNKNGSKIEEDPHIRSAALLPVLIHNSKKGPP 723
Db 478 -----EERPSTEPQVTKSGKVA----- 496
QY 724 ROAKVAIHCHAFPSKETOPAQIPEPLHSLDPSNLEHLITPLVTIGHIALADPOFA 783
Db 497 -----SSSKTK-----PTVP----- 506
QY 784 PWSKVATFIYKDLMLNDLPCKTKTKLMDPEEVSPEPTMYKIQAIKMAVRLMLKNNH 843
Db 507 -----PSKSTST-----ETKVAKOSEKRYVG-----SDNA 531
QY 844 SKSGSTLRLTLTILHSDGDLTEQGIKSPDKMSRLAAGSAIVLAQEPCHETITLQ 903
Db 532 QES-----TKPEKKKPGRGAI----- 550
QY 904 YOLCALINDECYOVFAQKLHKLRLPLPYMAICALCANDPYKERRAHARQCLV 963
Db 551 -----DE-----ESLHTSSGDNEKP-----AVSGKLASKSKKEAKQTYE 585
QY 964 KNIN---VRREYIKQHAHVSEKLSLEPEYVVPYITLHLLADPDYVYQDIEQLDVKEC 1020
Db 586 ESPSNNTKRKRSLSGGKASGESLVG-----SRIKV----- 615
QY 1021 LMFVLEILMAKNENNSHAFIRKMYENIKOTDAQGPDPAKMKNEKLYTQCDVAMNITMS 1080
Db 616 -WMPMD-----QAYYKGVVSY-----DAKKKHLVITDDGQELIYLN 654
QY 1081 TTYLSLESPKDPVLPARFTQPDKNFSNTKNYLPPEMKSEFTPGPKPTTNVLGAVKPLSS 1140
Db 655 QKWS----- 662
QY 1141 AGKQSQTSSRMETVSNASSNSPSPGRIGRLDSSSEMDHSENDYTMSSPLPKKSKD 1200
Db 663 S-ELSDDEADQGTQEDBASTVSGAGSSKAK-----ATPASKSK 703
QY 1201 -RDDSDLVSELEKPRGRKTPPYEOEKLGMDLTLVLOKPK-----GSOSRRKG 1253
Db 704 TSQDDKTASKSKDSKASREBEASSESE-----EEDPKTYGKSSSSSK-- 751
QY 1254 HTASESDQWPEEKRLLEDILENEQNSPPKRGKRGRRPPKPLGGGTPPEPTMTKTSK 1313
Db 752 -----DI-----SSVSKSGSKASSK-----KKEPSKAT-- 777
QY 1314 GSKKSGP-----PAPEEEEBEROSGN-TEQKSKSKOHVRRRAOQRAESPSSAISTQ 1368
Db 778 SSKSGSPGVKSPVASKTGKAKSGSASTPASKAKESASESESEETPREPE---PATK 833
QY 1369 STPOKRGRRPKT 1381
Db 834 AKSGKSGQSGSKS 846

RESULT 11
T30336
nuclear/mitotic apparatus protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30336
R:Merdes, A.; Ramyar, K.; Vecchio, J.D.; Cleveland, D.W.
Cell 87, 447-458, 1996
A:Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assem
A:Reference number: 820828; MIMD:97053784; PMID:8698198
A:Accession: T30336
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2253 <MER>

A:Cross-references: EMBL:Y07624; NID:g1514670; PIDN:CAA68905.1; PID:g1514671
C:Genetics:
A:Gene: NuMA
Query Match 3.2%; Score 231.5; DB 2; Length 2253;
Best local similarity 19.2%; Pred. No. 0.0046;
Matches 329; Conservative 248; Mismatches 535; Indels 601; Gaps 80;
QY 7 RTNDGKITYPGCVKISDKISKEEVRRKMYVTKFFMDQDSE----- 50
Db 272 RENQTELOEPKELELRK--NESLMIRDTLKOCDMKADKLLERKNDOLAEENGEL 329
QY 51 -----EKKELYL-----NTAL 61
Db 330 SKYVDLSNRLAQLOALYETTEBOELISNMQOKONOLESTLSCAVGEEKYLEHNIL 389
QY 62 H-----LASDFLKHGKDVRLVACLADIFRIVAPAYTSPDKLDF-----MTI 110
Db 390 QGKISMLMD-QLKEMGITDMPETGDMGIDLK-----DDLKQELAVLNTQCLSL 438
QY 111 TROLKGLDPTS-----POFNRYFYLENTAWKSNICPELDS--NEIFTOYRTL 161
Db 439 KEQIHOMEEKSTAEVEMDAOKSR-----FESEKGOLOEIVTNLQSTL 481
QY 162 FSVI-----NNGHOKVHHMVDLMSIIEGDTVSOEL-LPTVLV-----N 202
Db 482 SEITFOKERLDNEAQAQEH-----LMQITTLKLEISKLKSSLVHNEBELKGIH 532
QY 203 LVPAHKNLNKAQYDLAKAL--LKRTAQAE-----PYITTFNOVILMKTGISDLS 253
Db 533 KVEERNEKNQLENFKMLGNLIGITQOLESKTEKVDYLRBOQOKILCERDSTLTINE 592
QY 254 H-----VFDLLELYNIDSHLLSVLPOLERKLKSNDEERLYVKKLAKKFG--A 302
Db 593 YKCKKDENSGLVNTKRYKTELODHQSTLSVIE---KLASEKEELASKQVODDLAKMIGLA 648
QY 303 K-----DSELASQNKPLMOCYLGFRFNDIHVIRLCEVCRFASHCILMHPDLAKDLYEKLVR 358
Db 649 KCQINDSENDOSK-----SNA-----TVESLKAQ 674
QY 359 SHDPEAIR-----HDVIVSYVPAKKDILLVNDHLNLFVRETLDRKWRVRKAMWG 411
Db 675 LSEQSQSLKYKKVYSNMEVSENSKLQDOLSVSESLRHLEH----- 719
QY 412 LAQITKTYALQSAKDA-----AKOIANIKDKLHIITYONSDIDLVLERIFA 460
Db 720 ---LEKRTKFAASLIDLKRIISHLEEMKLSERDEALH---NLDEERTACKTIES 771
QY 461 QYMWPHNLETPRMCLYLYATL--DLNAVVALNE-----MMKC-----ONIL 502
Db 772 QL---KHLE-BEYOKANESLOAKLKGSCALIKOREEBDELISKVYDINKAKGESOQIA 827
QY 503 RH--OVKDLIDLKOPKTD--SYKAIFSKVAVITRNLPDPKADDFKKFTQVLED-- 555
Db 828 QNSCHMQOTEELKTKTHDOVYQOLEGERSKVLMTAKKASETSQ--LEKINQLEGELSA 885
QY 556 -----DEKIR-----KQELYVSPTSCQAOECVREITKK 586
Db 886 ANACIKREAEKEKLVLSALHSAEIKLIAVOGESERLSHLETALSNAKODLDC--LAKE 942
QY 587 LGNPQKPINPLEMIKPL---LERIAPVH-----IDTESIALIKOVNS 628
Db 943 LSDEYKKAEEFAMKVLKEONSERIASLESGLKSLAVKKEKSESEKSEVHILKQ 1002
QY 629 IDGTADDEDEGV-----PTDOAIRA--GLELLKVLSTFHPISPHSSET 669
Db 1003 LDDSSQKHKEALQAKNIEIKOLINAKKATSLAIKSEMGALQKAAVD--THKSEPSALON 1061
QY 670 FES---LLA-----CLKMDEKVAEAL----- 689
Db 1062 ELSRSLDLALKEGEVERLNKEALROEIOOQOQITRLTLEETALAKDKVALQOKE 1121
QY 690 ---QIFKNMGSKIEEDPHIRSAALLPVLIHNSK-----GPRQAKVAIHCHAFSSK- 740

```

Db      1122 IKQOYATKGA--EKEMAKLKS-----VISEKSRICLEDIODIONOKRDLSCIOEHOOSKL 1175
QY      741 -ETQFQO-----IIFERLKH-----SLDSNLEHLTPVLTIG 771
Db      1176 GESQGLALADLEKKCKEOKELICERQAKAEAKTLASEKASVSEBREGIOALEIEIG 1235
QY      772 HIALAPD--QPAAPW-----KSWATFIVKDLMM-DRLPGKKT-----KLWP 814
Db      1236 KERQKACDLOQKOLSEMAVOEEKETELOALKKELFIKVOLEBOSQTSFTDSGEBALLYLS 1295
QY      815 DEVSPEYVYKICATMMVVRWILGKNNHKSSTLTLLITLHSDG-----LT 865
Db      1296 EADEROQALTEAKEQAEYOKEIEMKNKEVNSLOAEIKILISKVTNNEEVSVDPEQRILK 1355
QY      866 EOGKIKRPD--MSRLRLAGSAIVKLAEPCYHEIITLBOYCALAINDECYQVRVFA 923
Db      1356 ETSKSAKLEKMKOKLMELEASFKELLEKCAIDCCTTEQONLKGEA--DOQMAVDSLO 1413
QY      924 QKL-HKGLSRLRLPLEYMAICALCANDPYKERBAHROCLVKNINVRXYLQHAAYSEK 982
Db      1414 OKLSKAETNHTLOEQIOAMQKCAE--KEQICISLO--QNLKSNQSLSEFASLKH 1467
QY      983 LLSLLEFYVYPTIYH---LLAHPDYKVO-DIEOLK-DYKECLEMYLETILAKNENSH 1037
Db      1468 YOEILAEERDMOEKHQEBELLSHKLTTERFOALELEKAKEDMTETV--LLEKELHNOELQ 1525
QY      1038 AFIRKVAENIKOTKDAQPPDAKM--NEKLYTCDAVAMNIMSKSTYSLSPKDPVLP 1095
Db      1526 KPOSEMSYSLTQSHLOQVNSOLGANQISQISDQAKKLESEMSTLK-EQKHE----- 1579
QY      1096 REFQDPKNEFSNTKNYLPPEMKSFTPPGKPTTNVLGAVKKPLSSAGKOSQTSRMETV 1155
Db      1580 -----EMKTLRL-----QYEKTLREGNKQVETSLOETV 1609
QY      1156 SNASSSSNPSPGRIGRDLSDSEMDHSENEDYTMSPLPGKKSDKRDSDLVYSELEKPR 1215
Db      1610 -----TSKYDH-----VSKVYLKDQ 1624
QY      1216 GRKTPYTEQEEK-----LGMDLTK-LVQEQPKFSQSR--KRGTASEDEQAPPE 1267
Db      1625 -----KTPOEEKORLLLOVQELNKLQSOQEKITRQOQKLKOREGTHEADK---SH 1674
QY      1268 KRLKEDLLENEDQNSPPKKGKRGRRPKPLGGSTPKPEEPMKTSKSKSKSGPPADEE 1327
Db      1675 KRYLE--LLEQLQEQVAVEHYKAKOMEK-----AKVHYDAKKKKON---QE 1714
QY      1328 EEEEROSGNTQSKSKSKOHRYVSRRAQOARSP 1360
Db      1715 LSEELQSHIKQOEHLSEN-----ADLKAESQ 1742

RESULT 12
S28261
centromere protein E - human
N:Alternate names: centromere 312k protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; PMID:93024922; PMID:1406971
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop

```

```

F:7-335/Domain: kinesin motor domain homology <KMOY>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Region: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match      3.28; Score 229; DB 1; Length 2663;
Best Local Similarity 18.24; Pred. No. 0.0076;
Matches 275; Conservative 268; Mismatches 537; Indels 434; Gaps 64;

QY      17 PGYKEISDKRSKEEMVRRLKMYKVTFTMDQDSE-----EKETLYNLALHASP 66
Db      333 PYNEVS--TDALLKRYR--KETMDLKQLEBVSLETRAQAMEKD--QLAQLEEK 363
QY      67 FFLKHGKDVRLVLACIADIFRIYAPAYTSPDKLIDF-METROLKGL-EDTKSPQ 124
Db      384 DLLQK-----YONEKENLTRLVYSSSLTQOELKAKR 417
QY      125 FNRIFYLLENIAVAVKSYNICEFELEDSNEITQILYRILFSYINNGHOKYMMHAYDLMSI 184
Db      418 KRRVYLCQKIRKKNSN-----YAOQFNIPN-ITTKHKISINLLREI 461
QY      185 ---ICEGDVYSOELDPTVL-VNLVPAHKNLNKOAYDLAKALKRTAQALPEYITTFENQV 240
Db      462 DESVCSSEVDFSTVTLDTLSEITEMNPAIKLLNQENIE----- 497
QY      241 LMLGTSISDLSBEHVEDLILELYNIDSLLSVLPQLEFKKSNDDEERLQVYKLLAKMF 300
Db      498 -----SELNLSRADYDNLV-----DYEQLRTEKEEMELTKRKNDLDEFEALERTK-- 545
QY      301 GANDSELASQNKPLMOCYGLRFPNDIHPVRLIECVKFRASHCLMMHPLAKRLTEYLAVRSH 360
Db      546 --KDOEQ-----LHEISMLKLVNHRVYNO 571
QY      361 DPEALRHQVIVSIVTAAKKDILVNDHLNPFRETRIDKRWVRKREAMGLAQIKYA 420
Db      572 DLNEL-----SSKVELREKEDQIKLQEIYDSK--LEIKKIDLS-----YS 613
QY      421 LOSAAGKDAKQIAMIKDKLLHITYONSIDRLIVERIFAQVWPHNLLETTERKCLYYL 480
Db      614 LESIEDPKQKQFLFDEAY-----ALDKARESAFLRESEULEKEKMKELATY 661
QY      481 YATLDLNAVAKALEMKKQMLRHOYDLDLKKOPTDASVVAIFSKVAVIR----- 534
Db      662 YKQME-----NDIQLOYSQLEAKKMQVDLE-----LQSAFELITKLSLDGKV 708
QY      535 -----NLDPGRKAQDFMKKFTQVLEDEDEKIRKQLEVLVSPSCQKQAEQCVREITKLG 588
Db      709 PKDLCLNLELEKITTDLQKELNKEVEENALBREV-ILLS--ELKSLPSEVRLKREI- 763
QY      589 NPKOPTNPFLEMIKFLERLAPVNIHDES-IALIKQVNSIDGTADDEDEGVPTQAT- 646
Db      764 ---QDKSEELHIITSEKDLFEVYVHKSRSVOGLEIGTKDLDATQSNYKSTDEQ 820
QY      647 -----RAGLELLKVLVSFTHISFSAETFEESLACLMD----- 680
Db      821 NFKTLHMFQKTKYKMLLENERMAQETIVNSK-----EAQKFBSSGALATKELSYKT 872
QY      681 ---DEKVAEALQIFKNKTSKIEDPFIHSAALLPVLNHS-----K 719
Db      873 QELQEKREVOER--NEMQOLKEQLENRDSPLQTVREKRTLLTEKLOQTLLEVYKTLQ 920
QY      720 KGRPRQAKVAILCHALFSSKEFOFOAIFERPLKSLDPSN-----LEHLTPVLTIGHA 774
Db      931 KDDLKQLOESLQI-----ERDLKSDIHITVNMNIDTOBOLRNALLESKLOHOETITWL- 983
QY      775 LLAIPDQAAPAKSVAVATFIYKDLMDRLRGKTKTKLWVDEVSPEPTVQIAIKMVR 834
Db      984 -----KSKISEVSRNLHM-----EERTGETKDEFO-----Q 1010
QY      835 WLIGMKNNHKSQSTLTLLITLHSDGLTEQKISK--PDMSRLRLAAGSAIVKLAQ- 891
Db      1011 KMGIDKKQDLEKNT-QTLTADYKNDIEILQQRKLFSLQEKNEQLQOMLESYIAEKQL 1069

```

QY 892 -----EPGYHEITLLEOYALCALAINDECYOVROVFAOKLHKSRLPLEYMACALC 946
 Db 1070 KTDKLENEMTINENOELRL-----LGDELKQOQELVNOEKHNAIK---EGELSTQCRL 1122
 QY 947 A--KDPVKERRAHAROCILVKNINVRREYLKOHAAVSE-----KILSLPEVVPYTI 996
 Db 1123 AEVEEKLKEKSOOLQOOLLNVOEEMSEMOKKINIEINLKNELKNELTLEHMETBRL 1182
 QY 997 HLHMH-PPDYVKYODI--DQLKDVKECLMFLVEILMAKNENNSHAFIRKRVENINOTKAO 1054
 Db 1183 ELAOKLKNENVEEYKSTKERVVLE---LQKSFETERDLRGVIREATEATGLQTKDEL 1237
 QY 1055 GPDDAKMNEKLYTVCDAVAMNIMSKSTYSLSEPKDPVLPAPFQOPDKNSNTKNYPP 1114
 Db 1238 KIAHILHKEHGETIDELRRSV--SEKT-----AQIINTQLEKSHTK--LQE 1280
 QY 1115 EMKSFETPPKPKTTNVGLAVKPLSSACKQOTKS--SRMETSVASSSSNPSPGRITG 1172
 Db 1281 EIP-----VLHEEQELLPNVKVSETOETMNELELTPQSTTKDSTLTARI-- 1326
 QY 1173 RLDSSEMDHSE-NEDYTMSSPLPGKSKDKRDDSDLVSELEKPKRGKRTPYTE----- 1224
 Db 1327 -----EMERLRLNKEFQESOEIKSLTKERDNLKTIKALEVKKHDKLEHIRETLAKIOE 1381
 QY 1225 -----QEEKLGM---DULTKLVQEO---KPK-----GSQSRKRGH-----TA 1256
 Db 1382 SOSKQOESLNKKEKDNETKTVSEMEQFKPKDSALRIEIMLGLSKLQSHDEMSVA 1441
 QY 1257 SESDEQO-----WPEERLKEDI-----LENEDQNSPPKKGKGRPP--KPLGGCTP 1302
 Db 1442 KKKDQLQRLQEVLOSQSDQLEKNIKEIYAKHLEETEELKVAHCCLKEQDEETINELRVNLS 1501
 QY 1303 KEETMTKSK-----GSKKSGSPRAPPEEESE---EHQSGNTQKSKSKQHRVSRRAQO 1354
 Db 1502 EKEEISTIQKQFALNDKLOKNKIOETIYEKEEQNLINIKOISEVOENVMLKQFKERRAKD 1561
 QY 1355 RA-ESPESASIEST 1367
 Db 1562 SAQOIESKMELELT 1575
 RESULT 13
 S46157
 R1F1 protein - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YBR1743; protein YBR275C; RAP1-interacting factor
 C:Species: Saccharomyces cerevisiae
 C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
 R:Accession: S46157; S46156; S44537; S25704; S39131
 R:Brandt, T.; Christiansen, C.; Holmstrom, K.; Holmstrom, K.; J. Ballese, T.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S46157
 A:Accession: S46157
 A:Molecule type: DNA
 A:Residues: 1-1096

 A:Cross-references: EMBL:Z36144; MIPS:YBR275C
 A:Experimental source: strain S288C
 R:Holmstrom, K.; Brandt, T.; J. Ballese, T.
 Yeast 10(Suppl.A), S47-S62, 1994
 A:Title: The sequence of a 32420 bp segment located on the right arm of chromosome II fr
 A:Reference number: S44537; MUID:94378722; PMID:8091861
 A:Accession: S44537
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1096 <HOL>
 A:Cross-references: EMBL:X76053; MID:9600025; PID:9429120

A:Experimental source: strain S288C
 R:Hardy, C.F.J.; Sussel, L.; Shore, D.
 Genes Dev. 6, 801-814, 1992
 A:Title: A RAP1-interacting protein involved in transcriptional silencing and telomer
 A:Reference number: S25704; MUID:92249772; PMID:1577274
 A:Accession: S25704
 A:Molecule type: DNA
 A:Residues: 1-579, 'S', 581-731, 'A', 733-1916

 A:Cross-references: EMBL:X66501; MID:94327; PID:CAA47121.1; PID:94328
 C:Genetics: SGD
 A:Gene: R1F1
 A:Cross-references: SGD:S0000479; MIPS:YBR275C
 A:Map position: 2R
 C:Superfamily: Saccharomyces cerevisiae R1F1 protein
 C:Keywords: transmembrane protein
 F:263-279/Domain: transmembrane #status predicted <TM1>
 F:296-312/Domain: transmembrane #status predicted <TM2>
 F:438-454/Domain: transmembrane #status predicted <TM3>
 F:663-679/Domain: transmembrane #status predicted <TM4>
 F:788-804/Domain: transmembrane #status predicted <TM5>
 F:905-921/Domain: transmembrane #status predicted <TM6>
 Query Match 3.2%; Score 227; DB 2; Length 1916;
 Best Local Similarity 17.4%; Pred. No. 0.0061;
 Matches 316; Conservative 267; Mismatches 612; Indels 624; Gaps 80;
 QY 3 HSKTRTNDGKITTY-----PPGYKEISDKI-----SKEMVRLKRV--- 38
 Db 183 YKETRVNFRALVELVWVGSEHGLVDNESYSEKEIEEGGLILROSEEDYVARREVYATF 242
 QY 39 -----VTFPMDDSEBEKELYLNLA-----LHLASDFLKHGKGVRLVLAACL 84
 Db 243 NNILPILTTKRVNVEVDKFNILVNIIEIETICPHQIADDTLLSSSEKKNPVYI--- 298
 QY 85 ADIFRIYAPAPYTS-----PD---KIKDIF-----MFTTR 112
 Db 299 -----RLVYQIVRFPSALIMSNFKIVKWLKRRDVLNKLVIKRVTTGALRNENSKIIITA 354
 QY 113 QLKGLDTSKSPQFNRYFLLENT--AWKSYVICPELEDSNIEFTQVLTLESVYNNHNO 171
 Db 355 QVSEFLRDEK---FQTFPLSNEIEKPIISTFEIMEINSHNLIYEKL---LLIRGLS 405
 QY 172 KVHMHVNDLMS-----IICEGDVSOELDDYLVNVPAAKNLNKQAYDLAKALL 222
 Db 406 KYPKLMIEYTVSMPLGEVLPRIITIGDELYSKILLITSLVVL-----ELL 450
 QY 223 KRTQAIEPYITTFEYNOVLMLGKTSISDSLSEHVEDLILEY--NID---SHLLSVLP 275
 Db 451 KKCLDPVDEHERIY--QCIMLSPVCETIPEKFLSKPLNSYDSANLKDVTIGHLITQOIK 508
 QY 276 QLEFKLSNDNEERLQV-VKLLAKMF--GAKDSELASONKPLM-----QCYLGRPNDIHV 327
 Db 509 --NYIVVKNDRKIMADMILMTGMTGLLYDSGKRVYDLSKSNKVPEDLNLCFINN---HP 562
 QY 328 PIRLECYK---FASHCL-----MNHPLDAKLDTLELK-----VRSHDPEAIRHDV 370
 Db 563 KTRLMSTIKVARIITTYCICIKTSQKNQENKSLSLRLPRPMTLPYVNDPSAREGCIITHL 622
 QY 371 IVSIVTA--AKKDIL-----LVNDHLNLFVBERTLDRKRVYR-----KEA 408
 Db 623 LGVVYVTAFTSNKNSLTDWFLFMDHLITPIYEDVVFYKDSIHLQNVLFVHLILGKNA 682
 QY 409 MMGLAQIYKKY-----ALDASAAGKDAK---QIAMIK---DKLHNIYONSIDDRILV 455
 Db 683 DVALERRYKKHHPMSYIASGVKLIKISSLPPI--IKREYDKIMKVVEQT-----V 733
 QY 456 ERIQAQVVPNHEETTEREMKCLVLYATLIDNAVAKALNEMKCONLRLRHQKDLIDLIIKO 515
 Db 734 EVASINNVNLADHLITS-----LKHLPED----- 757
 QY 516 PKTDASVKAIFSKVAVITRNLPDQKADQEFKKKFTQVL-----EDDEKIRKOLE 564
 Db 758 RKDQTHLESFSSLLTKVTQNNKDPPIPRDFEAGVTSFVYVTFDLFLRKNDSSLVNPNIQ 817


```

QY 565 VLVSPTCSCKQAEVCYREITTKLGNPKOPTNPELEIKFL-----ERIA 609
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 818 ISKVGISQGNMTLDELKDVIRKARN---ETSEFLIEKLELDCKTEVYAQWVGSTLL 874
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 610 PVHIDIESALIKOVNKSIDGDADEDEGVPDQAIRAGLELLKTLKTLSPHPSISAEI 669
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 875 PRISIRERFQSLANNYK-----VPENSIENFLDICKLTSF--PVNL----- 915
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 670 FESLACLMDDEKVAEALQIFKNTGSKIEEDF-----PHISALLPVLHKS 718
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 916 FTLLHVSMSNNNFYIQSYVSKNE-NKLNVDLITLTLSPGNELSELGILLPFL---- 970
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 719 KKGPPQAKY---AHCIA-----TESKKE- 741
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 971 -----RRKFMIDLEKCHISNPMLNSIPDLSDDLKLPKRSASYFANIKLFCSEQ 1025
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 742 -----TOFAOIFELPKSLDPSNLE--HLTPLTITGHIALLPDQFA 782
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1026 LTLVRLKLGQOLEQNLONFSEIENVLQNASD--SELEKSEILIREL--HLANADIE-- 1079
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 783 APKSWVATFIYKDLIMN--DRLPGKTKTLWVDEEVSPEYWKIQAIKMVRYMLLGMK 840
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1080 -PLFSGSLNFCIKNNNADHDERCGNMTSEVLF--KISPELLKILTYK----- 1125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 841 NNHSGKSTLRLTLTHSDGD---LIEGKISPDMSRLKLAGSALVLAQEPCTH 896
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1126 --EKPNGLLAIVIEKIENGDDDIILELEKIIIOKE-----IOILEKLEP--- 1170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 897 EITTEYOOLALAIIDECYQ---VROVFAQKHLKGLSRLLPLEYMAICLCKADP- 950
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1171 ----LAVFLNPNVSSMOKHKSSTNMLRELVLITLKPISRSAAKFFSMILSIDPPNP 1226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 951 -----VKERRAHAROCVKNINVRREYKQ---HAAVSEKLSL 986
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1227 YQIDMVNLLIDLKSHNRKFKDKRTY--NATLKTIG---KWIOESGVYHOGSSKEIEA 1281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 987 LPE---YV-----VPYTHLHLPDYKQVODIEOLKDYKEC 1020
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1282 IPDTKSMYIPCEGSEKNLSNLOKRVSDIOVPAT--QGKMEPPSSIQIISQISAKD--SDS 1339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1021 LMFVLEIILAKNNENNSHA-----FIRKAVENIKOTK-----D 1052
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1340 ISLKNALINSSQOESHANRSRSDIDEETLEEVONESIREDQOMKSTOLDKNAVANSNIC 1399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1053 AOGPDDAKNNKLYTVCAVAMNIMSSTYTESLEBKPVLPAARFETPDKNESNTKNYL 1112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1400 STKSDDEVLYTE--LHESIDTQOSEVNAVYOPLEVLTSELKAVTNSIKTPDHVNVSDNPL 1458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1113 PREMKSFTRPKPKT-----TNVLGAVNKPFLSA----- 1141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1459 KRPSKETPTSENKRSKGHEMTMDVLYSEQAVSPSSDVICTNKSIANESSIALRNSIK 1518
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1142 -----GKOSOTKSRMETVSNASSSNPSSPGRIGRGLDSS----- 1178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1519 VETNCNENSLANTLDDOQITIKEDKGQVEHYQROBENQDSNMKINSKSTODNIAQYKS 1578
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1179 -----MDHSENEEDYTMASSPLPEKSKDRDSDLVSE---LEKPRGKRTKPYTEOE 1227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1579 VKKARPNNGENNDCVNY-----EQASPVANEVPGDGIQIPSGITLLNSKQTE 1628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1228 KLGMDLTK-----LVQOKPKGSQRKRGHTASSEDQOMPREKRLKEDILEN--- 1277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1629 KSKVDDLRSDEDEHGTVAQOEKIQVGAINSRNKNNDMDSTPIQGTSEESREVVMTPEGIT 1688
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1278 ---EDQONSPPKKGKGRPKPLGG-----TKPEETPMKTSKSGSKSKSG 1320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1689 VRLBDSGTCELNKNLKG---PLKGDKANINDDFVPEENVYDEBGLFSMEHVAVSKETG 1744
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1321 PPAPEEEEEEE-----ROSGNTEOKSKSKOHARSRRAOQAPAESPESS 1362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1745 -----LEQGPYADVLSVLEIRIPINSLKMGSKQIKELKAKKRQORNLMPDPDPK 1799
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 1363 AIESTOSTPOKGRGPKST 1381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1800 T-ENTNINAQNGIDTVPKT 1817
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
T41581
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_rev15003 03-Dec-1999 #text_change 20-Jun-2000
C/Accession: T41581
R/Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z22002
A:Accession: T41581
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4717 <MUR>
A:Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08
A:Experimental source: strain 972h(-)
C:Genetics:
A:Gene: SPDB:SPCC737.08
A:Map position: 3
C:Superfamily: Saccharomyces probable membrane protein YLR106c

Query Match 3.1%; Score 223.5; DB 2; Length 4717;
Best Local Similarity 18.4%; Pred. No. 0.03;
Matches 290; Conservative 249; Mismatches 572; Indels 469; Gaps 71;

QY 24 DKISKEMVRRLKNVYKTNMDQDSEEEKLYNLALHLASPFLLKHGKQVRLVACC 83
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2935 DDVS--NLIEQIKV-----LDLNDISREIYFLASRLCSF--KQHSSDH-----S 2977
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 LADIFRIYAPAYTSPDKLIDFMETITROLKQLEDTSPOENRYEYLLENTAMVX---S 140
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2978 LANSFVLANEF-YINNAIKQ-----KELEIEEKNRLRYQREPNPKN-DYLVFIN 3029
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 YNICELEBSNEL-----FTQYKTLFVYINNGHOKYAHMNVDLMSI-----I 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3030 YDDEVPEVEPEVIEKRFLOLQFAFWSLYNEIYSEKNNVJPLEQIMTGSYLAKKIV 3089
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 CEGDTVSOELDVLNVLPAHKNLKNQAV-----DLAKALLKR 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3090 KNDMDIASSEFDIVSYVLMGVKSTNERQYTWTPPVYTFSDPSPSAIEYRDLIKIVESR 3149
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 TQAQIEPYITTFENQVLMKTSISDLS-----EHVFIDILEL-----YNI 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3150 AISLINMPENFVLRGKIDALINLSPSPIAEYLSKLERVFHLLSEMEKLASREYSL 3209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 DSHULL---SVLPQLEPKLSNDNERLOYVKLAKMFGA-----KQSELA 308
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3210 ANEMDILKRRKIIDWRKEFELSNMNNLKLLEYKLSERVYPRLYSIILOFILKPFENSKPT 3269
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 SONKPLMOCYLGFR--NDIHVPIRLCECVKFAHCLMHPDLAKOLTEYLVKVRSHDPEAR 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3270 KQNLCSASITIVQFTIDLTVGERQDLCK---CLLSFSQHAASL-----RIG 3312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 HDVIVSIVTAAKDKILLVN-----DHLNPFVERTLDKRWKVRKE 407
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3313 HGI-----DAMLNLIYHFEEQFLSKVSAIHTQKQSLNSIKERILLMSWMDT-- 3360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 408 AMMGLAQIYKKAIVAGSAGKDAKQIAWKIKDLHLIYYONSIDRLIVERIFAQYVPHN 467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3361 -----NVYALKESAKKSHAELF-----KVLHRYRE-----VLQPVSSYLSQKH 3400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 468 LETTERMKCLYIXATLIDNAVAKLMDM--KONILRHQV---DLIDLKQPTDAS 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3401 WD-----SLDTEENNSA---MMVAKKVNLSPSYIEEMQTEIMKLV--PVRFNS 3443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 VKAIFSKVMVITRLNLDPPCK-----AQDFMKFTQYVLEDE-----KIRKQ 562
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3444 TPTTLKMLMTLFLANVEKPGSTFTNNMVSNIITDARELMKLTPEITINDNLSLKHLSRKH 3503
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Qy	563	LEVJVSPTCSKQAGCVREITTKLGNKQPTNPFLPMKTLKLEIAVNHDTESIALI	622
Db	3504	L-LTLETFTKTLKAFGLQYRV--KAGIEENLSN-----LRNLAVIPPEPYSISIEKVD	3554
Qy	623	KQVNSKSID-----GTADEDECEVPDQAIKAGLETLLKVLSPFHPIPSHSAA-----T	669
Db	3555	RSLKSKSDPIKFPOTLAGHQNIDLSVP---EYQKCVGL-----FNSMLSLQGEKAAOLVE	3606
Qy	670	FESLLIACIK--MDDEKVAEAAIQIFKNTGSKIEEDFPHIRSALEPVLHHKSKKGPQAK	727
Db	3607	FTNELALKKNYSEVGVNGSPLESFPNNS-----SFNVVSS--LGYHDFENRA---QAV	3655
Qy	728	YAIICHAIFSSKETQFOAQIEPRLKSIDSPENLEHLITPLVTIGHIALLADQEPAPKMS	787
Db	3656	SMLOIYAIIVIOK-----HSSISPT---ASFOSIGH---ELSRFA-----	3688
Qy	788	WVATFIYKDLIMNRLPG-----KKTIKIWPDEEV	818
Db	3690	-----DILSNKLPSPSIPPLVYASADKVSSIRDOQKGINDLIEYCRKKREL--PELSY	3739
Qy	819	SPETMVKIQAIKMMVWRLMLGKNKNSKSGSTLRLITLILHSDGDLTPOGKISKPDMSRL	878
Db	3740	CFKHLVSLQSLKSTISRQVDTLND-----BFLNMNVNLPFSLLSISIEATFNMTFE	3793
Qy	879	R-LAAGSAIYKLAOEPCYHEIITTEQYQALAINDECYQYRVQFAOKLHGLSLRLP-	936
Db	3794	KEIAETSFIEIMSS--CFSKVLRAPNLFKFOGMSKLSLKEKIRSSVDMSCQMLMLFLPV	3851
Qy	937	-----LEY-----MALCALCADPVKERRAHANQCLVK	964
Db	3852	CEQITINIAESVLDTFINVHNSNLDLSKITLFLFVAVANGFCSPLDPOEGSNSGE----	3907
Qy	965	NINVRREYLKQAAV-----SEKLLSLPEYVVPYTIHLHADDPYVQDIE--QLKDV	1011
Db	3908	-----LESQGTGSGCVGAEDITNTLNE--DDDEBELANEDTANQSDLDSEAREL	3956
Qy	1018	KECLMFPVL-ELIMAKKENNSHAFTIRKVAENIKQTKDAQG---PDDA--KMKKELYTCDV	1071
Db	3957	ESDMNGVATKQSVSENESENSDS-----EEBNODIDEVNDIPDLISNLEIKLM-----	4004
Qy	1072	AMNIMSKSTYYSLESPKDPVLPARFPTQPKNFSNTEKNYLPPEKKSFTTGKPRTTNVL	1131
Db	4005	-----DEP-----NEEDLLETCKS-----NEQ	4022
Qy	1132	GAVVKPLSSAGKQSQTKS---SMETVSNASSSSPSPGRIGKRL--DSEMDHSENEED	1186
Db	4023	SAANNESDLYSKEDDNKALIEDKQOEKDEEMSDQVAGIDEIQDPIQENNSQPRENEED	4083
Qy	1187	YTMSSP-----LPGKKSDDKSDSLVRSLEKPRGKRKTPVTEQBEKLGMDITJLVQEOK	1242
Db	4083	H-LDLPEDLKIDKEGVGKSDSDIEMDMFAADEKKEADADEKDEP--MODFEDPLEENN	4139
Qy	1243	PKGQSRSKRRHTASSEDQGWPEPEKRLKEDILENEDQNSPPKKGKRGPRPKPLGGCTP	1300
Db	4140	TLDEIDIQDDPESDLAIDEDEKA--NEDGFEENVQENESTEDGVASDELEQ----GEVP	4192
Qy	1303	KEEPTMTKTSKSGSKKSGPPAPEEVEEEROSGNTQOKSKSKHNRVSRRAQOARSPRESS	1362
Db	4193	EDQALIDNHKPKMDAKSTFASAADEBENDMKXIVGENEHELGEEDGAALSGVVRNGTADGEFS	4255
Qy	1363	AIESTQSTPOKGRGRPKTP	1382
Db	4253	SABOVQ-----KGEDTSTP	4266

RESULT 15
T18296
Myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T18296
R:Gillien, N.
Submitted to the EMBL Data Library, February 1997

A:Reference number: 218865
A:Accession: T18296
A:Status: preliminary; translated from GB/EMBL/DDJB
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDs:AAB48065.1
C:Genetics:
A:Gene: mhca
F:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 3.1%; Score 223; DB 2; Length 2139;
Best Local Similarity 19.6%; Pred. No. 0.011;
Matches 292; Conservative 233; Mismatches 570; Indels 396; Gaps 62.

QY 2 AHSKRTNDGKI-TYPGVKEISDISEEAWR-LKNVVKTFMMDSDSEKELYNL 59
 | : || : || : || : || : || : || : || : || : || : || : ||
DB 894 ATAKGELEAKQDLLEDKISELSKSLAAELDKLNLTIE---NLEEKELKETIDL 950

QY 60 ALHLASDFELKHKGKDVRLVACCLADIFRIYAP-----EAPTSPDKLDMFT 111
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 951 KGDKLDS---KLKGGDELEVETELNSQINTLNATYDKKDQTIAEQESIDEKEDDTKLK 100

QY 112 ROLKGLDPTKSPOENRRFYLLLENIAW'-----KSYNICFELEDSNEIFTQLRYLF 162
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1008 GDIKLLEEKKD-----LEDDRADVSA TKDDIANKINKITIECEDAKDELTAKEOLE 1060

QY 163 SVINNGHN-----OKVHHHMYDMSIIIC-----EGPTVASOELLDPVLNVLPAN 207

DB 1061 DEENNKKDITNELNOOTJLKLGTEKSLAAQAATAATKASDERPOTLSQNLNEKLT'---- 1115

QY 208 KNLNAQAVDLAKALKLRTAQAITEPYITTFPNVOLMGKTSISDLSSEHFDDLLELYNIDS 267
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1116 KNLGTTKADLEK-----KISGLKQDYEDL-----EDDK 1143

QY 268 HLLLSVLPOLEFKLKSNDNEERLOYVKLLAKMFAGKDSLASQNKPLMOCTYIGRPNDIHV 327
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1144 NKIEGDLENNAQRIKIELDDE---ITK-----GADVSQYLQKQEEYESQIAKM----- 1188

QY 328 PIRLECVPFASHCMLNHHPDLAKDLLEYLVKSHRDEEALRHDIYSITYAAKCOLILAND 387
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1189 -----OEKEKAIGNVKKKEKTIKEKELETQS- 1215

QY 388 HLINFVRERTLDKRMVRVERKAMMGLAOIYKKY-AQASAAGKDAKOIMIKDLHIYQ 446
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1216 ----LOBKDETEVEKDEDAKKKKEILEKKAQOEKENVESKSTENDK----- 1262

QY 447 NSIDRLLVERIFAQYWPHNLETTERRKCCLYYIYATLDLAVALNMKMCQNLLRHQV 506
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1263 KILED-----NLKDTQ-----KLDMDTADNEKLKAKA 1290

QY 507 KDLLDLIKOPKTDAVKAIFFSKVMYITRNLPDPCAODPFMKFTOVLEDEDEIKROLEYL 566
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1291 KDLEQLNLEVO- DNHKAV-----ADAELLNKKKASQSDLESLKLEBAL 1335

QY 567 VSPTCSCQAEGCVREITFKLGNPKOPTNPFLFMIKFLLERIARPAHIFTESIALIKOVN 626
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1336 T-----KANSVSESKKDSNEKAALS---EELDAQNEKIKNTIQADLRATADLOQAN 1387

QY 627 KSIDGTADEDEGVPTDAIRAGLELKVLSFTNPISSHSAEFESLACLMDKDEKVAE 686
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1386 EKKAIVEAQRDLVANDKRKMTTLEIRKARDEENT--KYVENYEKVLKRREADLEE-AN 1444

QY 687 AALOIFKNTGSKIETEDFPHISALLPVLIHNSKKRPPOAKYAINHIA-----IPSSKET 742
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1442 ENLDEKKDRMNKEQV-----KLEBELKETRKDMAAIAEDSIFETARKQ 1488

QY 743 QFAQIEEPRLHSIDSNSLEH--LTPLVTHIGHIALARPAPKSWAFYIVKDLLAN 800
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1489 SDADL-EELINTIVE---EHDEVAKLAT--QITRLTDNOSAEBE-----LNELSK 1533

QY 801 DRLPCKRTTKTLWPDDEV-----SPETVWKI-----QAIRMWYRWLGM 839

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 20:20:49 : Search time 236 Seconds

(without alignments)
891.814 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193
Sequence: 1 MAHSTRNDRKIRTPGVK.....QKGRGRPSKTPSPQPKNV 1391

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 566894 seqs, 151307093 residues

Total number of hits satisfying chosen parameters: 566894

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	758.5	10.5	363	US-09-925-302-800	Sequence 800, App
2	231.5	3.2	2835	US-09-885-535-4	Sequence 4, App1
3	225	3.1	1743	US-09-882-227-624	Sequence 624, App
4	206	2.9	2243	US-10-118-513A-12	Sequence 12, App1
5	197.5	2.7	1530	US-10-118-513A-6	Sequence 6, App1
6	197.5	2.7	2266	US-10-118-513A-14	Sequence 14, App1
7	189	2.6	3907	US-10-171-311-2	Sequence 2, App1
8	189	2.6	3925	US-10-171-311-6	Sequence 6, App1
9	186.5	2.6	2649	US-10-205-219-169	Sequence 169, App
10	186	2.6	1979	US-10-205-823-419	Sequence 419, App
11	185.5	2.6	2437	US-09-815-242-5834	Sequence 5834, App
12	185.5	2.6	6281	US-09-815-242-12996	Sequence 12996, A
13	185	2.6	780	US-09-770-689A-5	Sequence 5, App1
14	184.5	2.6	1814	US-10-128-714-3388	Sequence 3388, App
15	184	2.6	676	US-09-823-187-24	Sequence 24, App1

16	183.5	2.5	1400	US-09-764-1176-7	Sequence 7, App1
17	183	2.5	670	US-09-823-187-86	Sequence 86, App1
18	183	2.5	671	US-09-946-374-308	Sequence 308, App
19	183	2.5	671	US-09-823-187-85	Sequence 85, App1
20	183	2.5	671	US-10-015-387A-308	Sequence 308, App
21	183	2.5	671	US-10-006-130A-308	Sequence 308, App
22	183	2.5	671	US-10-199-672-346	Sequence 346, App
23	183	2.5	671	US-10-006-112A-308	Sequence 308, App
24	183	2.5	671	US-10-187-749-346	Sequence 346, App
25	183	2.5	671	US-10-194-457-346	Sequence 346, App
26	183	2.5	671	US-10-184-642-346	Sequence 346, App
27	183	2.5	671	US-10-196-747-346	Sequence 346, App
28	183	2.5	671	US-10-015-392A-308	Sequence 308, App
29	183	2.5	671	US-10-017-253A-308	Sequence 308, App
30	183	2.5	671	US-10-173-689-346	Sequence 346, App
31	183	2.5	671	US-10-173-690-346	Sequence 346, App
32	183	2.5	671	US-10-173-691-346	Sequence 346, App
33	183	2.5	671	US-10-173-692-346	Sequence 346, App
34	183	2.5	671	US-10-173-694-346	Sequence 346, App
35	183	2.5	671	US-10-173-698-346	Sequence 346, App
36	183	2.5	671	US-10-173-699-346	Sequence 346, App
37	183	2.5	671	US-10-173-707-346	Sequence 346, App
38	183	2.5	671	US-10-174-569-346	Sequence 346, App
39	183	2.5	671	US-10-174-583-346	Sequence 346, App
40	183	2.5	671	US-10-174-587-346	Sequence 346, App
41	183	2.5	671	US-10-174-589-346	Sequence 346, App
42	183	2.5	671	US-10-174-591-346	Sequence 346, App
43	183	2.5	671	US-10-175-736-346	Sequence 346, App
44	183	2.5	671	US-10-175-742-346	Sequence 346, App
45	183	2.5	671	US-10-175-744-346	Sequence 346, App

ALIGNMENTS

RESULT 1
US-09-925-302-800
Sequence 800, Application US/09925302
Patent No. US2002004941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 800
LENGTH: 363
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (358)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-302-800

Query Match 10.5%; Score 758.5; DB 9; Length 363;
Best Local Similarity 45.0%; Pred. No. 5e-42;
Matches 172; Conservative 56; Mismatches 105; Indels 49; Gaps 11;
OY NINVRREYIKOAAVSEKILSLPEYVYPTHTLHPDDYKVDIOLDKVRECLMEV 1024
DB 1 NISTREIKONPAATEKILSLPEYVYPTHTLHPDDYKVDIOLDKVRECLMEV 60
OY LEIMAKNENNSHAFIRKVENIKOTKAOGPDAAKMKNEKLYTVCDAVMNIMTSKSTYS 1084
DB 61 LEVIMTKNENNSHAFIRKVENIKOTKAOGPDAAKMKNEKLYTVCDAVMNIMTSKSTYS 120


```

QY 1186 DVTM-----SSPLPGKSDKRDSDLVRS---ELKPRGRKKTPTVEQEXL----- 1229
| : : : : : | : : : : : | : : : : : |
Db 1569 ATVOOQLEMERLNHOALLKLLKATYTESDKTTPVODPLMELKLIMDSLEERTINROHKLEGAL 1628
| : : : : : | : : : : : | : : : : : |
QY 1230 -----GMDDLTK-----LYOQOKPKGS-----OR 1248
| : : : : : | : : : : : | : : : : : |
Db 1629 LALGOFQHALDELLAMLTHTEGLLSEOKRPVGPRAIELEIAKHVLONDVLAHOSTVFA 1688
| : : : : : | : : : : : | : : : : : |
QY 1249 SKRGHTASED-----EQQWPEERKLKEDILENEDQNSP-----PKK 1287
| : : : : : | : : : : : | : : : : : |
Db 1689 VKNAGNDLIESAGEBASLONKLEVLNQRW-----QNVLEKTEQRKQOOLDGALROAK 1741
| : : : : : | : : : : : | : : : : : |
QY 1288 GKRGK-----PKPLGK-----GTPKE-----EPTMKP-SKK 1313
| : : : : : | : : : : : | : : : : : |
Db 1742 GRHGEIEDLQOMLTTERILLASKPLGSLPETAKEOLNVHMEYCAFAFEKETYSLMOK 1801
| : : : : : | : : : : : | : : : : : |
QY 1314 GSKKSGPAPPEEEROSGNTQSKSKOHVSRRAQORAES 1358
| : : : : : | : : : : : | : : : : : |
Db 1802 GQOMLARCP-KSAETINIDODINNKKEMESVETKLNERTKLEEA 1845
| : : : : : | : : : : : | : : : : : |

RESULT 3
US-09-882-227-624
; Sequence 624, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; FILE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 1743
; TYPE: prt
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 876
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-882-227-624

Query Match 3.1%; Score 225; DB 12; Length 1743;
Best Local Similarity 18.4%; Pred. No. 8.2e-06;
Matches 285; Conservative 217; Mismatches 523; Indels 520; Gaps 71;

```

```

Db 467 -----MKKTLER-----YNDCLKNAT----- 483
| : : : : : | : : : : : | : : : : : |
QY 277 LEFKLSNDNEERLOVYKLLAMPGAKDSLSQNKPLMOCYLGRENIDHVRILECYKF 336
| : : : : : | : : : : : | : : : : : |
Db 484 -----EERIKCIDLI-----KIDENL-----KKSIL-----LNQOKVQALDCLKN 518
| : : : : : | : : : : : | : : : : : |
QY 337 A-----SHC-LMHP-----DLAKULTEYKRSHPREAIRHVDIVSYVPA 378
| : : : : : | : : : : : | : : : : : |
Db 519 AKTDEERNCLTLINDEPEIREKFRKELELOKLEQV-----KDKIKNAKTEA 565
| : : : : : | : : : : : | : : : : : |
QY 379 KKDILL--VNDHLNFRERTLD-----KRMVRKFAAMGLQIYKKYALQ----- 422
| : : : : : | : : : : : | : : : : : |
Db 566 EKNKCLGSKRAIERLKQALDCLKNKATDBERNCLKNITQDLOKELLAMSVKAYND 625
| : : : : : | : : : : : | : : : : : |
QY 423 --SAAGKDAKQIAWIKDLIHTIYQNSIDRLVERIFAQYVPHNET--TERMKCLY 479
| : : : : : | : : : : : | : : : : : |
Db 626 CVSKARNEKEQEC--EKLLPPEARKKLEQVL-----DCLKNKATDBERKKCLMD 674
| : : : : : | : : : : : | : : : : : |
QY 480 LYATL--DLMAYKALN-----EMKQONL-----RHQVDDL 512
| : : : : : | : : : : : | : : : : : |
Db 675 LPRKDLOSDLIAKESLKAAYKDCVSOAKTEAKKECEKLLPPEARKKLEBEAKSVAYLDC 734
| : : : : : | : : : : : | : : : : : |
QY 513 IKOPKTDAVKAIFSVAVITRNLPDPGAODPMKFTQVLED-----DEKIRQOLEV 566
| : : : : : | : : : : : | : : : : : |
Db 735 VSQAKTEAKKEC-EKLLT-----PEARKKLEAKSVKAYLDCVSRARNEKEKECEK 788
| : : : : : | : : : : : | : : : : : |
QY 567 VSPSCQKQAEGBVEITKILGNPKOPTNPLEMIKFLERIPAPHIDTESIALIKOVN 626
| : : : : : | : : : : : | : : : : : |
Db 789 LTPPEAK-KLEQOALDCLKNKATDBERKKCLDLPKIDQKV-----LAKESVKAAYLDCV 843
| : : : : : | : : : : : | : : : : : |
QY 627 KSIDGTADDEGCVLPDOAIRAGLELLAYLSTHPISRSATFESLACLKMDKVAE 686
| : : : : : | : : : : : | : : : : : |
Db 844 QAKTEAKKECEKLLTPKA-RKLEBAK-----KSVAXLDC-----SQ 882
| : : : : : | : : : : : | : : : : : |
QY 687 AALQIFKNTGSKIEEDPHIRGALLPVLIHKKSGKGPPOAKYAI-----HCIAIFSKET 742
| : : : : : | : : : : : | : : : : : |
Db 883 AKTEAKKECEKLLT--PEARKL-----EPAKSVKAYLDCVSRARNEKEK 927
| : : : : : | : : : : : | : : : : : |
QY 743 QFAQIFELPKSLDPSNLEHLITPVTIGHIALAPDQFAABKWSVATFIVKLLMDR 802
| : : : : : | : : : : : | : : : : : |
Db 928 K-----ECEKLLTP-----EAKKLEEQ-----VLDCCLKNK 954
| : : : : : | : : : : : | : : : : : |
QY 803 LFGKTKTLKVPD--EVSPEPMVKIOAKMMVRLGKKNHSGSTLTLLILHS 860
| : : : : : | : : : : : | : : : : : |
Db 955 TEADK--KRCVADLPKIDQKVLAK-ESVKAYLDCVSRARNEKEK----- 997
| : : : : : | : : : : : | : : : : : |
QY 861 DGDILTEQGISKPDMSRLBLAAGSAIVKLAQPCYHEITTEYOYCALAINDCYOVNO 920
| : : : : : | : : : : : | : : : : : |
Db 998 -----ECERLTPPEARKKLEBAKESL--KAYRDCLSQARNEERACEKLLPPE--ARK 1047
| : : : : : | : : : : : | : : : : : |
QY 921 VFAOKLHKL-----SRLRLPLEYMAICALCAKDPVERAHRAROCVLK--NINVRE 971
| : : : : : | : : : : : | : : : : : |
Db 1048 LLEQEVKYSIKAYLDCVSRARNEKEKECEKLLTPEAKFLAKOYLNCLEKNGNEERKA 1107
| : : : : : | : : : : : | : : : : : |
QY 972 YLKOHAASEKLLSLPEYVDPYTHLLAHDPDYKVDIEQLKDKVCEKLEVLITLAK 1031
| : : : : : | : : : : : | : : : : : |
Db 1108 CLKN-----LPKIDQ--NILAK-----ESLKAYKDCV-----SQAR 1137
| : : : : : | : : : : : | : : : : : |
QY 1033 NENNSHAF-----IRKVENIKOTKQAGDDAKMNEKLYTVQDVAMNIIIMSTIYS 1084
| : : : : : | : : : : : | : : : : : |
Db 1138 NEERRACERKLLTPEARKLLE-----QEVKSVKAYLDCVSRARNEKEKECEK 1186
| : : : : : | : : : : : | : : : : : |
QY 1085 LESPKDPLVPARFTOPDKNFEN--TKNYLP-----PEMKSFPTPKPPTT 1128
| : : : : : | : : : : : | : : : : : |
Db 1187 LITPEARKFLAKELQOKKAIKIDCLKNADPNBRAIMKCLDLSDEKIKYIOERAKVA 1246
| : : : : : | : : : : : | : : : : : |
QY 1129 NVLGAVNKPLSAGOSOTKSRMETVSNASSSSPSSBGRITKGLDSE--MDHSEN--E 1185
| : : : : : | : : : : : | : : : : : |
Db 1247 ADCLAMARTDEEKRCOMLYSDLIOENKRTONQONOLSKTERLHOQASECINDDDPTD 1306
| : : : : : | : : : : : | : : : : : |
QY 1186 DVTMSPLPGKKSDDR-----DSDSLVRSLEKPR-----GRKKTPTVEQEXL 1230
| : : : : : | : : : : : | : : : : : |

```

Db 1307 OEALIEGLEGISDERALLIGIKROADVEDLIYSDLRNRKTFEDNMAAKGYELPMDEKNG 1366
Qy 1231 MD-----DLKIVOEOKPKGORSRRKRGHTASESD--EQWPEERKLEIDLENEDE 1280
Db 1367 GDATINATNDADKISDNI-----VASIEDIAKQYETETITDKMLE----- 1412
Qy 1281 QNSPPKGGKRGPRPLGGCTPKKEEPTWTKTSKSGSKKSGPPADEEBEEBESRGNTQOK 1340
Db 1413 -----AKLAKALGNG--KKDDDEKESKSTAEKAKENNKIDKVAETAKNISRIA 1460
Qy 1341 SKSKOHVSRRAQORAE--SPSSAIESTOSTPOK-----GRCRS 1379
Db 1461 LKNNKESGEEFVDENGNPIDDKKAEKODETSVPKQAFIGSDPT 1505
RESULT 4
US-10-118-513A-12
Sequence 12, Application US/10118513A
Publication No. US20030039953A1
GENERAL INFORMATION:
APPLICANT: Taga, Tetsuya
APPLICANT: Kimura, Naoki
TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
FILE REFERENCE: 06501-107US1
CURRENT APPLICATION NUMBER: US/10/118, 513A
PRIOR FILING DATE: 2002-04-08
PRIOR APPLICATION NUMBER: PCT/JP00/05756
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: JP 11-288738
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: JP 11-288739
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: JP 2000-123721
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 2243
TYPE: PRT
ORGANISM: Mus musculus
US-10-118-513A-12
Query Match 2.9%, Score 206; DB 15; Length 2243;
Best Local Similarity 18.0%; Pred. No. 0.00021;
Matches 290; Conservative 218; Mismatches 509; Indels 590; Gaps 74;
Qy 195 LLDIVLVNLPANHNNKQAVDLAKALKRTQAQ-IEBYITTF----- 236
Db 526 LLSRLIDIDOPSSIQEBOLEALISAAIQSSIGLGLTGYIKRWIIIEOPNSAANLRYLE 585
Qy 237 --FNQVIMLGKTSISDISEHVED--LILELYNIDS-----HLISVLPOL-EFKSN- 284
Db 586 WTKWKV-VLTKEEDRCLVPLFDGSCRFIDPQTIQSIOQCHLLSNLSTVASCAMEOG 644
Qy 285 -----DNERLOYVKILAMFGAKSELASQNKPLMOCYLGRNDIHVPILCEYKA 337
Db 645 ITERGLVDSLNRKHNVTQLCOY-----AHMVLMFCHSGL-----LREGDDLALH 689
Qy 338 SHCLMNPDLAKDLLEYLKVASHDPEARIRHDVIVSYITAAKDDILVHDLNLPVREPT 397
Db 690 SRLRYNTPVIONYTTSRQKSEKSPRGKMHND-----CLMIDGLVSOQDGE- 735
Qy 398 LDKRMRYRKEAMMG-----LAQIY-----KKYA-----LOSAGK----- 427
Db 736 VEKIMK-RDECGTGRYPASTHALDIDYLDNITEASHALITTYILLDIMKSFNKDTTP 794
Qy 428 ----DAKQAMIKDKL-----HIYQNSIDRLLVERIFAQ-----YVPHNLE 469
Db 795 IESEPTAFATSMGVQKIVQGFMLLDHNDYENGLD--LLFHPYAKPASWQSHKIIIEAFMS 852
Qy 470 TTEEMKCLYLYL-----ATDLNAV-----KALNEMMKC-QNLLRQVQDL--- 510
Db 853 QGEHQALRYLQTMKPTVSSSNEVILHLTVLLFNRCVNEAMNLLRONSNRVNIIEELLKHA 912

Qy 511 -----DLIKQKRTDAS--YKAIFSXYVITRNLDPGKAQDPM----- 546
Db 913 YEVCQENGIMEDLKLFTPTNTEQECIVKPLQSSSV-----ENHEFLVHHQGRAN 963
Qy 547 ----KRTQYLED-----DEKIRKOLEVLVSPICSCQKABGVREITTKLGNKPOKPTNP 596
Db 964 YISALKINOLIKNNIMSDRDRPLRE-----RSYTR-----NS 995
Qy 597 FLEM-----KFLERLAPVHIDESI-----SALIKOVNSIDGADDED 637
Db 996 ILDOYKILPRVQKLAVERAKPYHLSTSVFHEVSRPKPLSAPK---KAITGTVLTBS 1052
Qy 638 EGVV-----TDQAIRAGLELLK-----VLSFTPIPSHASETESLIACL 677
Db 1053 TFIENVLSKIGEVWASHBPRNGVSLFNSPKTEQSPVYVSHPPH----- 1096
Qy 678 KMDKQVAEALQI-FKNTGSKTIEDPPHRSALLPVLHKKSGKGPQAKYAIHCIH-- 734
Db 1097 ----ELPEAFVGPISTISQRI-----SRLDLVHV-----PVPQSQCLEFTQOS 1138
Qy 735 ----AFSSKETQFAQIEPPLHKSIDPSNLLEHLIRPLV--IGHITALAPDOFAAPWK 786
Db 1139 PTRSPLCLLSSSLPLSSQFKRPHONTSRPSLELLLEPLLYKAKSIALSAT-----S 1191
Qy 787 SWAFETIVKDLAMDRLPGKTKTL---WVPDEVSPEMVKIQAIKMYRWLLGKNN 842
Db 1192 SGFAEFPPPSILRS-----GFRPTPLASPSLSPGSLIAPPFRVKETRISFME--GNMTH 1244
Qy 843 HSKSGT---STLRLLTITLHSDG-----DL-----TE 866
Db 1245 WTRATDTRNTKAVSVSTFHCQGPAPETEMKTSDKNTYRPLDPAKQKVAASLATH 1304
Qy 867 QGKISKPDMSRLRLAASAIYKLAQEPCEYHETLEQYOCAL-----A 910
Db 1305 SGRLEKLDVSKEDSTASTRSDQISLE--YHDAPSEDLBGAVFVSPKPASSITELTINST 1362
Qy 911 INDECYOVROYF-----AQLHKLGISRLRLPLE---YMAICALC-- 946
Db 1363 LQTRERDNDKAFKSEGTSPYPYKQIGTGDAVAEAFSELSRHD--VEERABASFGVSVCEG 1421
Qy 947 -----AKDPYERRAHAR 959
Db 1422 ETSTNSKTSYLDGIVPIESRTSLTADHKESVANTVADYESSGSTSKCAPYSESLGQ 1481
Qy 960 QCLVKNINVRREYKQAAVSEKLSLPLP---YVVP---YTHLHAPDQYKVDIE 1012
Db 1535 E-BAKNLSPDILVPLGAKEKLEYNLSTEQFCOLDPODKSABCDAAEVDGELF----- 1586
Qy 1073 MNTIMSKSTIYLSLSPDPVLPARFTQDPKDNFNTNTNLYLPPEKKSFTTGKPKTAVLG 1132
Db 1587 --VAOSNFTLILEEGE-----EAESDSAAAMNLPKSTK----- 1619
Qy 1133 AVNKPILSSAGOSQOTKSRMETVGNASSSN-----PSSGRIGRLDSSMDHSENEYD 1187
Db 1620 --EKPVCYRPHNDRERTDLPASAVTADQESIKVETLPYVEPQVVALAENLIDYIKOTRS 1677
Qy 1188 TMSGPLPKSKSKDROSDLVRSLELEKPRGKRTPVTEOEKL-----GMDLTL----- 1235
Db 1678 KEATPVAAAGEAGDEDAVIYASKAHSNLNSTPKYKPEAPRAETVNTSQSDQWVSRLTL 1737
Qy 1236 -----KLVOEQKPKGQSRKRGHASE-----SDEQQWPE-----E 1267
Db 1738 RQHALSLNTVSEQEPASAVATPKKRTKIKETPESSERTOSDLKVAENOLITAOONPPAPR 1797
Qy 1268 KRLKEDI-----LENDE--QNSPPKGGKRGPRPKPLGGGT--KEEPTMTSKKS 1315
Db 1798 RRRKDVSOQGLTPSSGAVEPEPEPOGTPGRRLRTPQPEAPEETPRTKVLSSVYKGT 1857

Db 1359 RSSRTRSKAILLPDLSEPNNEPLFSPASEVPRKAKAK--KIEVPAQLKEIV 1408

RESULT 6

US-10-118-513a-14

Sequence 14, Application US/10118513A

Publication No. US2003003995A1

GENERAL INFORMATION:

APPLICANT: Taga, Tetsuya

APPLICANT: Kimura, Naoki

TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS

FILE REFERENCE: 06501-107051

CURRENT APPLICATION NUMBER: US/10/118, 513A

CURRENT FILING DATE: 2002-04-08

PRIOR APPLICATION NUMBER: PCT/JP00/05756

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: JP 11-288738

PRIOR FILING DATE: 1999-10-08

PRIOR APPLICATION NUMBER: JP 11-288739

PRIOR FILING DATE: 1999-10-08

PRIOR APPLICATION NUMBER: JP 2000-123721

PRIOR FILING DATE: 2000-04-19

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 2266

TYPE: PRT

ORGANISM: Homo sapiens

US-10-118-513a-14

Query Match 2.7%; Score 197.5; DB 15; Length 2266;
Best Local Similarity 18.2%; Pred. No. 0.00078;
Matches 293; Conservative 223; Mismatches 603; Indels 493; Gaps 68;

QY 56 YLMLAHLMSDFL-----KHFGDVALLYAACADIFRIYAPARTSPDKLDTF 107
DB 750 YPPASHAVLMDLTLGGVTEAARHS-----ITTYLLDITMYSF-PKKTDTPIESPTVF 802
QY 108 MFTTRQIKGLGEDPKSPQFNRYFYLLNIAMVKSYNICFELEDSENIPTQLYRTLESYINN 167
DB 803 AISMGOVKLIQG-----FWLIDHNDYESGLDILFHATAKAPLSWOSKTIIOAFMSQ 853
QY 168 GHNOQVHMWVNDLMSSTICEGDTVSQELDTVLV-----NLVPAHKNLQAYDLA 218
DB 854 GEHROA-LRYIQTMKPFVSSGNDVIILH--TVLLFNRCVEANWFLEOHGN----- 901
QY 219 KALKRTQAIEPIYITTFNQVLMIGKTSISLSEHFEDLILEYXINDSLLSVLQLE 278
DB 902 -----RLNTEELKHMYEVCQEGMLDEL----- 926
QY 279 FKLSNDNEERLQVVKLLAMFGAKDSELSQNKPLMOCYLGRFNDIHPRIEVCYFAS 338
DB 927 -KLPFTDTBEGC-LVNFLOSSASVQNHFF-----LLVHHLQRAH--YVF----- 966
QY 339 HCLMNHDPDLAKDILEYLVKR-SHPDEAIRHDVIVSIVTAKKDIILLVNDHLNPFERT 397
DB 967 -----ALKTINQTLKINVMNDRPRLRE-----RSIARNSI 996
QY 398 LDKRMVYRKRAMMGIAQIYKYVALQSAAGDAKQIAMIWKLLHIYYONSIDRLLYER 457
DB 997 LDQYGI-----LPRVHRKLAIERAKP-----YHLSTSSVFRVSR 1032
QY 458 IFQAYWPHNLLETTERKCLYLYATLDLNAVAKALNEMKCONLNR----- 503
DB 1033 PKPLSAVPKQVTVGTVLTRSVFIN-----NVLSTIGEVMASKEPINSTPTFPNSKILEEPS 1087
QY 504 -----HOYKDLLDLIKQPTDASYKALF-----SKYMTI 532
DB 1088 PIYVSLPAPLPEAFEGTPTISKASQKISRLLDLVYGVPRPSCQSEPTIOQSSMKSPLYLV 1147
QY 533 TRNLDPBGKAQDMKFKFTOVLEDEKIRKQLEVLVYSTGCSQAEGCVREITTKLGNPKQ 592
DB 1148 SRSLPSSSQLKGSPOAISRASFE-----LHLEETPLV-VKRAKSLAMSVT----- 1190

QY 593 PTNPFLMIKIFLERI-----APVHIDTESIALIKQVY-KSIDGTADD 635
DB 1191 -TSGFSEFTPQSLRSTPRSTPLASPSRPGSRPQRLKTRISFVEDVHPKPIRGAAAD 1249
QY 636 EDEGV---PTDQIRAGLELLKVLSTPHRISFHSATFBSLLACLMDDEKVAEALQITF 692
DB 1250 SKLEVPFTPKCAVAPETEMPKSDRTTFSLSPKEHQ-----EMDEGSOLEKLDVS 1304
QY 693 K-NTGSKIEDPFIHSALLPVLHHSKGGPRPQAQVAIHCHIAITSSKETQFAQFEEL 751
DB 1305 KGNSSVSTSDETTLLEYQAP-----SPDLEIYVFTASKPKSSSTALTITNVTBQT 1355
QY 752 HKSIDPSNLEHTTP---LVITIGHI-----ALLAPDOFA-----APKSWVAAPFIVD 796
DB 1356 EKDGDVDVASEVTPSDLOKQMNLEDAETKDLVAAAFSELNHLSPVOGTAEASLCAPS 1415
QY 797 LLMNDRLPGK-KTTKILAMPDEEVSPTETWIKQAIKMMVWMLGAKNNHKS-----G 847
DB 1416 V-----YEGKIFQKSKVP---VLDEGLTSVETYPFAIR-----ANDKSMADVLGDDG 1461
QY 848 TSTRLRLTLTILSHDGLTEBQKISKPDMSRLRLAASATVLAQE-----PCYHETI 899
DB 1462 NSSLTITSEGIYSEIRLNQEVALNLKEDHEVEYVUKESYDLPEEKLPISDSPPTQIETH 1521
QY 900 TTEYOULCALAINDCYOVRQYFAQRLHKGSLRLPLEYMAI---CALCAKADPYKERR 955
DB 1522 VIEQEKLEA---QDSSEARNLISFNELYPS-GTLKIQYNFDITDQFCPLADKKDTAE-- 1575
QY 956 AHARQCLVKNIN-----VREYIKQHA--ANSEKL----- 983
DB 1576 -----CDIAEVDDELVAQSNFTLILEGEGEVEPDFASSDVLPRKAAVTAETEEKLVCSG 1630
QY 984 -----LSLPEYVY---PYTHLHAPDQVYKODIEQKDV----- 1017
DB 1631 ENDNHQIAMLPSAVISDQSKQVDTLPVPEPIKALAEMLDVLKDRSKETITSDTME 1690
QY 1018 ---KECLWFLYEITAMKNNENSHAF-----IRRVENIKQTKDA----- 1053
DB 1691 QSTHERITPLVQSINMCPTRLVKSAPFAQGTSTMTNMVSVQVDVSSKTRTGRIQIONVN 1750
QY 1054 -----QCPDDAKMNEKLYTVDVAMNIIIMS-KSTTYSIESKDYVLA 1095
DB 1751 VKSQAQESADAVATPRMFGOSVRRKTKRKAKEISEASENITSYVGLFEQNOOITPONSVTPR 1810
QY 1096 RFTTOPDKNPSNKNLPRPMKSFETPGKP-----KTTNVL--GAVNK-----PLSSAGQ 1144
DB 1811 KGRKKKEVNODILENTSSVDEIQITTGRESKRLKSSQLEPAVEETTKKEVAVSSYTKR 1870
QY 1145 SQTSSSR---METVS--NASSSSNPSPGRIKGRLLDSSEHSHSNEEDYTMSPLPGKS 1198
DB 1871 TPRIRKRSVENQESVELINDLKVSIVTSPSRMIRKLRSTNLDASENTG-----NKQD 1922
QY 1199 DKRDDSGLVASELEKPRGKKTPTVTEQEF-KLGMDDLTKIYQOKKRGSGRSRKRKHTS 1257
DB 1923 DKSSDQRLRIKHRRVARGREVSPSDVREDSNLSLSOLT--VQAEFDM--SAIPKRKRPRK 1979
QY 1258 ESDEQOQPEE---KRLKEDILNEDQONSPPKRGGRPRK-----PLGGG--TP 1302
DB 1980 IN-----PSHDVSKAKKEERSPKKKEADPSIRKRSTRNTPANSENVGCKPALGKSLYP 2034
QY 1303 KEPTPM-----KTSKSKSKKSGPAPAEDEEEROSGNTQORSK-----SKOH 1346
DB 2035 NEELSMVMSKKRLTKKTESQSKRSLHSVSEERTDEMTHKENDEERLILASTTKSS 2094
QY 1347 RVSRRAQORA-----ESPSSAIESTQOS--TPQKGRGRPSKTPSPSQPKKNV 1391
DB 2095 RSSRTRSKAILLPDLSEPNNEPLFSPASEVPRKAKAK--KIEVPAQLKEIV 2144

RESULT 7

US-10-171-311-2

Sequence 2, Application US/10171311

```
Publication No. US20030087270A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Glatf, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersth, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-035
CURRENT APPLICATION NUMBER: US/10/171,311
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 3907
TYPE: PRP
ORGANISM: Homo sapiens
US-10-171-311-2
```

Query Match 2.6%; Score 189; DB 15; Length 3907;
Best Local Similarity 18.5%; Pred. No. 0.0063;

Matches 283; Conservative 267; Mismatches 564; Indels 416; Gaps 73;

```
19 VKEISDK-----ISKSEENVRLKVVYKTFMDMDODSEEEKEL-----YLNLALHLASDF 68
2118 LKEKDKSELLISKEQLOIR-----DIOERNEIEKLEFRVRELEQLALVSADTF 2167
69 LKHPKDVRLVACCLADFRIRYAREPRTSPDKLDFMTTRTOLKGLIEDTKSPQFNRY 128
2168 QK-----VEDKKHFGAVEA--KPELSLEV-----DLQAEKRAIDRKEKI 2205
129 FYLEINIAWKSYNICFELEDSENEIFTOYRTLFVYINGHNOYMHMVAWDMSSIIICG 188
2206 TMLSEQLQGFRE-----ELENNKNE-----EYQOLHMQL----- 2233
189 DTVSELDLTVLVNLVPAHKNNLKOAYDLAKALKRTAOAIEPYTTF--FNQVLMIGKT 246
2234 -EIOKKESTTRLOELEOEENKLFKDDMEKLGIAIKESDAMSTODQHVLFCKFAOIIQOEK 2292
247 SISDSEHYFDLILBLVYNIDSHLLSVLPQLEFKLSNDN--EERLOVYKLLAKMFGA-- 302
2293 EIDQLEBOYTKL-----QOOLKITTDKVIIEKNNELRIDLETOIECLM 2335
303 KDSSELAOKPLMOCYLGRFNDIHPVIRLECYKFAASHCLMNHDPDLAKDLTEYLKVRSDP 362
2336 SDQECVKNR--EERIEQLNEVIEKIQOELANIQKTSMAHNSISE----- 2380
363 EEAIRH--DVIVSYTTAKKDLILVNDHL--NFRERTLD-----KRWVRK 406
2381 ADSLKHOLDVYIAEKIALEOQVEETNEMTEPKNVKLKTFNKMNOITLOELSKRERESV 2440
407 EAMMGLAOIYKKVALQSAAGKDAKOIMIKDKLHI-----YQON-----SIDR 452
2441 EKIQSIPENSVNAVIDHLSKDKPELEVLTEDALSLDENQTYFKSFEENGSGSITINLETR 2500
453 LLVERIFAQYVPHNLLETTERKCLYLYATLIDNAVALNEMMKQ--NLIRHQYKDLDD 511
2501 LL-----QLESTVSAK--DELFQCYQOIKDMQOQGFETEMIQKTIYNLOK 2545
512 LKQPKTASVKAT-----FSKMWVTTIRNL--DPGKAQDBMKKFTVOVLEDEDEKIRKOL 563
2546 IVEEVAALVSOJOLEAVOEVAKFCODNOTISEPERTN--IONLNLRIEDE--IGSDI 2601
```

```
564 EVLVSPITSCQAEGVREITKKLGNPKOPT-----NPLFEMIKFILERTIAPVHIPTESIS 619
2602 SAL--TLRISLEESQVEMHNSLILEKOVEIAEKNVLEKKEKLE----- 2645
620 ALIKOVNKSIDGT-----ADDEDEGVPTDQAIRAGLELLKLVLSFTPHISFSAET--FESL 673
2646 -----LOKLIEBNEKKOREKREKRSPOD-----VEVLK-----TTLEFHSNEESGFNE 2690
674 IACLKMDKEVA--EALQIIFKNTGSKIEED-----PPHRSALLPYLHNSKSGPPRO 725
2691 LEALRA--ESVATKELASYKEKAELQOELLVETKNTMSIQKDISQVADHDA-----E 2742
726 AKYALHCHIAIYSS--KEQFOQIIEPRLKSIDPSNLEHLITPLYTIGHIALADQORA 783
2743 AKEKSTIEKEDETEVOESKKACMEPRLPIKLSIASO-----TDGLKLISSNSO--- 2793
784 PKMSVAVPFIKDLMLNDRLPKTKTKLWVPDEVS-----PETMVKIOAKIMVAVRML 837
2794 -----TPOLVKNAGIQIMLOSECS-----EVTETIISOFTKIEKKQ--ELIRAEIL 2840
838 GKNHNSKSGTSTLR--LLTTLHSDGDLTEQOKISRPMSRLRLAAGSAIVKLAOEP 894
2841 DMESRHI--SETETIKREHYVAVOL-----LKECGTFLKAVIQLSKSGESSIPELAHSDA 2894
895 Y--HEITILEQ-----YOLCALAINDECYQVQVRAQIKHLSRLPL 937
2895 YQTRICSSDGSMDGQGIYTLHSGQFDIASGRGESEBSAIVDSPKKI--KELLR--AVHN 2952
938 EYMAICALC-----AKDP--VKERRAH-----AROCVKNINVRRE-----Y 972
2953 EEMQVLSLSTESPYSGEDHSIQVSEPIEKKAIINTISLKDITMKQLOREAVYDS 3012
973 LKQHAVS--EKLISLLEPVYVPYTHLHNDPY-----VKYOD 1010
3013 SOSHESFSDMRRELLLAQOVLEERSVILAAFPFELALGTTDAVGLNCLQIOG 3072
1011 IEOLKDVCELMFVLEIIMAKRNNNSHAFIRKMNENIQTDAQPD-----DAKMKK 1064
3073 VE--YQAMECLOKADRSLSEIQALHAOMNGRKITLKEQSEKPSQELLEYNIOKOS 3131
1065 LYTQVDVAMNIIIMSKSTYTSLESPKDPVLPAPRFTQOPKNSNFKNYLP-----P 1114
3132 QMLEQVLSLSSMKDQATLEQDLSSEKNNVAKLSE-----LAQTKLEETLAKQHKILK 3187
1115 EKSFTTPGKPKPTNVIGAVNKPPLSSAGKOSQTSRMETVSNASSSSNPSPGRIGRL 1174
3188 ELEAFRLVKKKTDEV--HLNDTLASEQKRSREIQMALE-----KERAKL 3231
1175 D--SSEMHSENEDTYMSSPLPGKKSDDRDDSDIVRSELEKPRGRKKTPTVTOEKLGMDD 1233
3232 GRSEERKDEEEDLKFSS-----LESOKORNLQNLNLLLEQOKOLLNES 3273
1234 LTKLVOEQKPKGSGRSRKRGHYAS--ESDEDOOMPKEKRLKE--DILENEDRONSPBK 1286
3274 QOKIESQMLVDAQDSEEGGNLLEQVLES-----EKVARIEMSSFLDERELHAQDQ 3327
1287 KGR--RGRPRPLAGGTPKPEBPTMTKSKGSKKSGPPAP--EBEEBEOSGNT--EQK 1340
3328 SSDGTGQSRPPLP-----SEDILKELOKOLEEHSRIVELNLETETKYKIDSLOTROOM 3380
1341 SKSKO--HRSRRAOORAESPESSAIESTROS 1369
3381 EKDRQVHKRTLOTQEDANTBGOCKMHELOS 3410
```

```
RESULT 8
US-10-171-311-6
Sequence 6, Application US/10171311
Publication No. US20030087270A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
```

APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Gialt, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersch, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: OF CERVICAL CANCER
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 6
LENGTH: 3925
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-311-6

Query Match 2.6%; Score 189; DB 15; Length 3925;
Best Local Similarity 18.5%; Pred. No. 0.0063;

Matches 283; Conservative 267; Mismatches 564; Indels 416; Gaps 73;

19 VKETSDK-----ISKEMVRLKMYVTGDMDDSESEKEL-----YLNLHLASDF 68
2118 LKEKTDCESELLSKEDLQ-----DIOERNEIEKLEFRVRELEQALVSDTF 2167
69 LKHGKQVRLVACCLADIFRIYAPARYTSPDKLQIFMIFITPOLGLGDTKSPQNR 128
2168 QK-----VDRKHFCAVEA---KPELSLEY-----QLQAEKRAIDRKEI 2205
129 FYLENIAMWKSYNICFELDSNEIFTQLYRTLFVINGNHOKVNHMYDLMSICEG 188
2206 TNLSEQLQEFRE-----ELEKNKNE-----EVQQLHQL----- 2233
189 DYSQELLDVVLVLAHKNLKNQAVDLAKALIKRPAQAIERYITTF--FNQVLMIGKT 246
2234 -EIQKKESTRLQELQENKLFKDKMEKLGIAIKESDAMSTODQHVLFQFAQIIOKEV 2292
247 SISLSEHVFDDILEYINIDSHLLSVLPQLEFKLGNND--BERLQVVKLAKMFGA-- 302
2293 EIQDLNQVTKL-----OOQLKITTDKVIIEKNELIRDLQLETOICLM 2335
303 KDSLEASONKPLMQCYLGRFNIDHVPTRLECVKFAASHCLNHPDLAKDLTEYKVRSDP 362
2336 SDOECVKNRR---EEETIEQLNEVIEKLOQELANIGKTSNMHSLSE----- 2380
363 EEAIRH--DYIVSVTPAAKDDILLVNDHL--NFVVERFLD-----KRMVYK 406
2381 ADSLKLQDLVYIAKLTALQEQVETANEMTFMKNVLTETNKNQLOLQELFSLKRESEV 2440
407 EAMMGLAQIKKYYALGSAAGKDAKJAMIKDKLLH-----YYON-----STDDR 452
2441 EKIOSTIPENSVNVAIDHLSKDKPELEVLTEDAKLSLENQTYKRSFEENGKSTITNETR 2500
453 LLVERIFQAYVBNHETTERMKCLYYLATLIDNAVKALNEMWKCQ-NILRHQVKDLD 511
2501 LL-----QESTVSAR---DLELTQCKYKQKJDMQEQGQFTEMLQKKTIVLQK 2545
512 LIRQPKTDASYKAI-----FSQVMYITRNLPDPKADDEMKKFTQVLEDEDEKIKQL 563
2546 IVEEKVAAALVSOIQLEAVQETAKFCODNOTISSEPERTN--IQLNLQREDE--LGSDI 2601
564 EVLVSPSCQKQAGCVREITTKKLGKNGKOPT---NFMILMIFLERIAPVNHDTESIS 619
2602 SAL---TLRISELESQVVENHTSLILEKEBOVEIAEKNVLEKKEKLE----- 2645

QY 620 ALIKQVKSIDGT---ADDEBQVPTDQIRAGLELLKVLSPTHPISFSAET--FESL 673
DB 2646 -----LQKLEGNENKQREKERRKSPD-----VEVLK-----TTTELFSNDESGFNE 2690
QY 674 LACIKMDEKVA--EALQITKKNQSKTEED-----FPPIRSALPVLHKKSKGPPRQ 725
DB 2691 LEALRA--ESVAARKAELASYEKAERKQLEBELVYKETNMTSLQDLQSVGRHLA-----E 2742
QY 726 AKVAIHCHIAIFSS--KETQFAQIFEPPLHSLDPSNLEHITPLVLTGHITALAPDOFAA 783
DB 2743 AKELSLILEKEDETEVOESKACMFEPDPLKLSKTSIQ-----TDGTLKISSNQ--- 2793
QY 784 PWSWVAFTYKDLMLNDRLPGKKTTLKWPDEEVS-----PETVYKIQAKIMYRWML 837
DB 2794 -----TPQILVKNAGIQINIQSESS-----EEVTEIIISQFTEKIKMQ--ELHAAEIL 2840
QY 838 GMRKNHKSQSTSLR---LTTLLHSDGDLTEQCKIKRPMKSRRLAAGSAIYKLAQEP 894
DB 2841 DMESRHI--SETEPLKREHVAVOL-----LKEBCGTLKAVIQCLRSKESGSIPELAHSDA 2894
QY 895 Y--HEITLQO-----YOLCALAINDECYQVQVFAQKLHKSRLPL 937
DB 2895 YQTRICSDSGSDMGOGITLTHSQGFDIASGEGESESATSPFKKI--KGLLR--AVHN 2952
QY 938 EYNAICALC-----AKDP--VKERRAH-----ARQLVKNINVRRE-----Y 972
DB 2953 EGMQVSLTESPSYDGEDHSIQGVSEPMLEBRKAYINTISSLADITKMQLOEAEVYDS 3012
QY 973 LKQHAAYS---EKLSLDEYVVPYTHILLANDPY-----VKVQD 1010
DB 3013 SOSHESFDMRGELLALQOVFLSEERSVLLAARTELTALGTDAVGLNCLBQRIQEOG 3072
QY 1011 IEOLKDVKECLMFLVLEITLMAKNENNSHAFIRKVENIKQTKDAQDP-----DAKMNK 1064
DB 3073 VE--YQAMECLQKADRSLSIEQIALHQAQNGKRTYKROESEKBSQELLEFTNIQOKS 3131
QY 1065 LYTVCDAVMMIINSKSTYLSLESPKDPVLPAREFTQDKNFSNTKYL--P 1114
DB 3132 QMLEQVETLSMKDRATELQOJLSEKMYVAELKSE---LAQTKLELETTLKAQNHK 3187
QY 1115 EMKSFTPGPKRTTNVGAIVNKPPLSSAGKQSQKSSMELVSNASSNSPGRIGRL 1174
DB 3188 ELEFAFLREVLDKDEV--HLINDTLASQOKSRELQWALE-----KEKAKL 3231
QY 1175 D-SEMDHSENEYTMSPPLPGKKSKDRDSDIYVRSLEKPRGKKTPTVQOEKLCMD 1233
DB 3232 GRSEERKELEDLKES-----LESQKRNQLNLLLEQOKOLNES 3273
QY 1234 LTRLYOBQKPKGSGQSRKRGHTAS---ESDEQWPEERKLE--DILENEQNSPPK 1286
DB 3274 QOKIESQRLMYDQOLSEQGRNLEQVLES-----EKVRIEMSTIDRERELHQAQ 3327
QY 1287 KGR---KGRPPKPLGGTTPKEEPTMKTSKSGSKKSSPPRP--EEEEEROGNT--EQK 1340
DB 3328 SSDGTGSRPPLP-----SEDLKLELQKOLEKHSRIVELLNETEKYKLSIQTRQOM 3380
QY 1341 SKSKQ--HRVSRRAQOARSPESSAIESTOS 1369
DB 3381 EKDRQVHRKTLQTBQEAQNTBQKKMHQLOS 3410

RESULT 9
US-10-205-219-169
Sequence 169, Application US/10205219
Publication No. US20030138803A1
GENERAL INFORMATION:
APPLICANT: Warner-Lambert Company
APPLICANT: Lee, Kevin
APPLICANT: Dixon, Alistair
APPLICANT: Brooksbank, Robert
APPLICANT: Pincock, Robert
TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain

```

: FILE REFERENCE: WL-A-018200
: CURRENT APPLICATION NUMBER: US/10/205,219
: CURRENT FILING DATE: 2002-07-24
: PRIOR APPLICATION NUMBER: GB 0118354.0
: PRIOR FILING DATE: 2001-07-27
: NUMBER OF SEQ ID NOS: 197
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 169
: LENGTH: 2649
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Pemphigoid antigen
US-10-205-219-169

Query Match      2.68; Score 186.5; DB 12; Length 2649;
Best Local Similarity 18.88; Pred. No. 0.0052;
Matches 257; Conservative 232; Mismatches 506; Indels 369; Gaps 69;

QY 193 OELDVLNVLPAHKNLKNKQAYDLAK-----ALTKPTAALIEPYT-----TEFNQVLM 242
DB 433 OELAEOLLEENHRA--RLNTEAYRAAMOTOMSWITLOLCOCVEOHIKENTAYEEFFMDA-- 488
QY 243 LKTSISDSEHVFDDILELYNIDSHLLSVLEPQEFKLKSNDEERLQYVVKLLAKMFGA 302
DB 489 ---KEATDYLRNLKDAIORKYSCDRSSSIHKLEDL-VQESMEKEELQYKSTIANLMGK 544
QY 303 KDS--ELASQNKPLMOCYLGREFNDIVPRL-----ECKVFSHCIMNHP 345
DB 545 AKTIILKPRNS---DCPL-----KTSIPKICDYKQIETIYKDECV-----LANN 591
QY 346 DLAKDLTEKLVKSHDPEEIRHDVIVSYTAKKRIILLVNDHLMFVREHRTDKRWRR 405
DB 592 HRAK-----WKVISPGNEAMWPSVCFVPPPKKEAVDLANRDEQYQNVLT--WH-- 641
QY 406 KEAMGLAQIYKKYALQSAAGDAKQAIAMIKDKL--LHIYONSIDRLIVERIFAQYM 463
DB 642 -FSHNMKSIVSWHYLINEIDRIRASNVASIKTLMLEGHOVLNSIQSR-----PEDF- 693
QY 464 VPHNLETTERMKCLYLYATLNLNAVKALNEMKCONLLRHQVKDILLDLIKOPKTASVK 523
DB 694 ---LEDSQESQ---VFSGSD-----ITOLEKEVNVCKOYQELKSAEREQESYV 739
QY 524 AIF-SKVAVYI-----TRNLDPGKAODFMKPFQVLEDDKIRKOLEVLVSP-- 569
DB 740 NLYISYVRNIRLREKCEBRLIRQITPRLERDDLHESVRITE--QELKLEERLDDIG 798
QY 570 --TCSCQ-----AEGCVREITTKLGNKOPTNPLEMIKFLLEIRIAPHI--DTESTI 618
DB 799 TITNKCEEFESQAAASSVPTLRSELNVVLQNNQVSSSTYIDKLTQVNLVLMKTAQA 858
QY 619 SALIKOVKNSIDGTADDEDEGVPTQDAIAGLELLKVLSTFHPISHSASEPESLLACJK 678
DB 859 EALVKLYETKL-----CEEAVYADK-----NNIENLISTLK 890
QY 679 MDEKVAEALQIFKNTGSKI-----EEDPHIRSAALLPYLHNSK----- 719
DB 891 QMRSEVDEKR-QVFALDELELOKAKAISDEMRKTYERDIDPDHWEKADQLVERQNVH 949
QY 720 -----KGPPOAKYAIIHCINAI-----FSSKETQAFQIEPPLKSLDPSMLEHLTP 766
DB 950 VOIDNRLRLEGIGKSLKYRDTYHPLDMDIQOVERTQKRIQGN-----QPEMSKILAQ 1004
QY 767 L-----VTIGHIALIAB--DQFAAPKMSWATFIVKQDLMNDRLPGKKTTKLVMPDEEVS 820
DB 1005 LNOOKLWSEIEIKOSKMDKCKYAEQYSGT--VKDYELQ-----TMYFRAMWDSQOKSP 1057
QY 821 EFMVKIO-AIKMNVRLJGMKNHNSKSGSTLRLLTTILHSODLLEQOKISKPDMSRLR 879
DB 1058 VRRRQSSADLIQFMDLRTY---TALVTLMQYIKFAGDSLK--RLEEEETKRC 1111
QY 880 LAAGSAIVLAOEPCHETII-----TLBOYOLCA-----LAINECVQ 917

```

```

DB 1112 -----FTSEHGAVSDDLQROKATVLENSKLTGKISELERNVAELKKOKSVEBELPK 1163
QY 918 VROVFOAKLHK-----GLSLRLPLEYMAICALCAD-----PKERRAHARQC- 961
DB 1164 VREAENELRKQORNVEDISLOKIRASE-----AKOYRELETTIVREKAEARELE 1215
QY 962 LVKNIVVRREYLKQHAASEKLSTL---LPEYVVPYTHILHMDPYVVOVDEOLKADVK 1018
DB 1216 RVROLITLAE--AKRAVAEENLINFNRQLENT--FTRRTL--EDHLRKDL-SINDLE 1267
QY 1019 ECLMFVLEITLMKNENNSH-----AFIRKVENIKOTKDAQDPD--DAKME 1063
DB 1268 QOKNKLMEEELRRKRDNEEELKLKIQEMKDLAQKQVAE--KOLKEKQKITELEARKITE 1325
QY 1064 KLYTCDVAMNITMSKSTYSLESPPQVLPARFFQPPDKNSNTNYLPPPEKSFPTG 1123
DB 1326 IOYTCRENAL-----PVCQ---IYQATSCRAVVG--HQOE-----HD 1357
QY 1124 KPRTNVLGAVNKPLSSAGKOSQTKSSRMETVSNASSSNPSGPRIKGRLOSSEMDHSE 1183
DB 1358 KQKAEELKQOQVDE-LTAANRKAQODKRELTYELNALQLEKTS--EKAARLLKDKIDETN 1414
QY 1184 NEDYTWSSPLPGKKSDDKSDLVNSELKPRGRKKTPTVEQEKLGMD--DLTKLVQBO 1241
DB 1415 NTLRLCKLEL-----ERKQDAEKGYSQLRELGRQNLQTTGKAEPAMQASDLKIKRMV 1469
QY 1242 K-----PKGSQSR---KRGHTASEDDQ-----QMPPEERL-----KEPI 1274
DB 1470 QLELESINHEKGLQREVDRIYRAHVAVAKNIQHLNSQIHSPDEKELRLQIQKRSKH 1529
QY 1275 LENEDQNSPPKKGKGRPPKPLGGCTPREEPLMKTSKSGKSSGPPAPEEEREROS 1334
DB 1530 LKQKQFKS-----HEQLQIWK--AEKENNDKQRLNELEKSN 1566
QY 1335 GNTQO-KSK---SKOHRVSRRAQRAESPESAIESTOSTPOK 1373
DB 1567 ECAEMLKQKVEELTRONNETKILMQRIQAESENIIVLEKQTIQOR 1610

RESULT 10
US-10-205-823-419
: Sequence 419, Application US/10205823
: Publication No. US20030108963A1
: GENERAL INFORMATION:
: APPLICANT: Schlegel, Robert
: APPLICANT: Monahan, John E.
: APPLICANT: Endege, Wilson O.
: APPLICANT: Gannavarapu, Manjula
: APPLICANT: Gorbacheva, Bella
: APPLICANT: Hoersch, Sebastian
: APPLICANT: Kamatkar, Shubhangi
: APPLICANT: Monsey, Angela M.
: APPLICANT: Glatl, Karen
: APPLICANT: Zhao, Xumei
: APPLICANT: Anderson, Dustin
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
: TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: FILE OF INVENTION: THERAPY OF PROSTATE CANCER
: FILE REFERENCE: MRI-044
: CURRENT APPLICATION NUMBER: US/10/205,823
: CURRENT FILING DATE: 2002-07-25
: PRIOR APPLICATION NUMBER: 60/307,982
: PRIOR FILING DATE: 2001-07-25
: PRIOR APPLICATION NUMBER: 60/314,356
: PRIOR FILING DATE: 2001-08-22
: PRIOR APPLICATION NUMBER: 60/325,020
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: 60/341,746
: PRIOR FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: 60/362,158
: PRIOR FILING DATE: 2002-03-05
: NUMBER OF SEQ ID NOS: 455
: SOFTWARE: FastSeq for Windows Version 4.0

```

SEQ ID NO 419
 LENGTH: 1979
 TYPE: PRF
 ORGANISM: Homo sapiens
 US-10-205-823-419

2.6%; Score 186; DB 15; Length 1979;

Query Match
 Best Local Similarity 19.1%; Pred. No. 0.0037;

Matches 290; Conservative 226; Mismatches 567; Indels 434; Gaps 65;

```

QY 16 PGVKEISDKISEEYVRLKVVKTMDQDSEEEKLY---LNLALHLASDFLKH 71
DB 362 PSAQKSDTMTKEKRIQASAVEEVRLOALDAENIMRLSSLNDNSLAD----- 416
QY 72 PGKVVRLVACCLADIRIYAPPEAPYTPSKLADIFPFITQLKGLDTPKSPQNRFFYL 131
DB 417 ---NLKLMR-----LEVEKEKSLSQEK-EELQMSLKL-----LNNEEV 454
QY 132 LENIAYKSVNYICELED-----SNEIFTQYRFLFSVIN--NGHNOKVHHMV 178
DB 455 IKSTRA-TRDISLDELHDLRLNLEAKQDELNQSISE-KETLIAIEIELDRONQATKHM 512
QY 179 ---DMSIICEGDTVSQELDTVLVNLVPAHKMLNQAVDLAKAL----- 221
DB 513 LIRQLSKQONEGDSIISKLODINDEKKRYHQ-LEDDKMDITKELDQKEKLIQSEVAL 571
QY 222 --LKRTQATPEYITTFPNOVLMGKTSIS-----DISEHVPDLIELVYIDSHLLSV 273
DB 572 NDHLTQKLEDKVENLVDQLNKSQESVNSIQENLELKEHIRONEEELSHIRRELMOQL 631
QY 274 LPQLEFKISNDNEERLQVVKLAKMFGAKDELSAQKPLMQCYLGHFNDIHPVIRLEC 333
DB 632 NQDSNSNPKDTLLKEREAEVNLKQNL-----SELEQLENL-----KVAEDVKNEN 679
QY 334 VKFASHCIMNHDIAKLDTLEYLKYRSHDEPAIRHDIYIVSVTAKKDILVNDHLLNFV 393
DB 680 EKLVIAC-----EDVRHQL-----ECLAGNNQL----- 703
QY 394 REPTLDKRMRYKREAMGGLAQYKKYALQSAAGDAKQALMIKDLILHIYQNSIDRL 453
DB 704 ---SLEKNTIVE-----TLKMEKGELEA-ELCWAKKRLL-----EEN 737
QY 454 LVERIFAQYVPHNLLETTERMKCLYUATLDLNAVALNEMKQCNLIRHQVLDLDLI 513
DB 738 KYEKTIEELSNARMLNLSALOEHENH---IKLNQKKDM-----ELAEKKNI 782
QY 514 KQPTDAS-----VKAFISVAVITRNLPRPGQAQDPMAKFTOVLDEDE 557
DB 783 EQMDTDHKEKTDVLSSLEQKQLTOLINKKEIFIEKLKERSKLOEELDYQALRKNE 842
QY 558 KIRQLRVLVSPQSCQAEQVREITKLGKPNQOPNPELMKFFLE---RIAPYRID 614
DB 843 ILRQITE-----EKDRSLGSMKEENNHHLOEELERLREEDSKTAPV-AD 884
QY 615 TESTISALIKOVNKSIDGTADEDEGVPTDQAIRAGLELLKVLSTFTHPSFASCTFESIL 674
DB 885 PKTLDV-----TELASVSQNLNTKEHLEEKH-- 914
QY 675 ACLKMDKVAEALQIFKNTGSKIIEEDPFRHSALLPVLHNKSKGPPROAKYALHCIH 734
DB 915 -----HOKIIEQONQSKMQLLOSLQEOKKE--MDEFRRQ---H 947
QY 735 AIFSSKETQAFQIEFPLHKSIDPSNLEHLIPRLVIGHIALADQAPAPKSWAVNTIV 794
DB 948 EQMNAHTH---QLF--LEKDEIKSLQKTIQITQLH----- 980
QY 795 KDLLMNDRLPGKTKTKLWVDEEVSPEPMVKAQIKMVVRLWLGKMN--NHSKSGSTL 851
DB 981 -----EENQDIOQNSDIFQET--KVQS-----LNIENSEKHDLSKATE 1019
QY 852 RLTLTIHSDGD-----LTEQ-GKISKPDMSRRL-----LAAGAIYKLAQEP 893
DB 1020 RLVVGIERELEIKLNLKKNISLTQLDQLSKDEVGKLTQITQOKDLEIQALHARISSTS 1079

```

```

QY 894 CYHEILTBEOYALCALAINDECYQVRFPAOKLHKGLSRRLPLEYMATICALADPKVE 953
DB 1080 HTQDVYTLQO-QLOAYMERE-----KVFVNLKERTENSHLKTETIKKMDIYA--AKE 1130
QY 954 RRAHARQCLVKNINVRREYLKONAAVSEKILSLPEYVVPYTHLLANDPYYKODIE- 1012
DB 1131 AALIKLQDENKKUJLSTRESSQD-----MRETIQNSLR---IIRKQIEI 1173
QY 1013 -QLKDYKECLMFEVLEIMAKNEN---NSHAFIRKMWKNIKOTDAQSPDDAKNEKLYTV 1068
DB 1174 DALSOQKQTLAVLQTSSTGNEAGVNSHQF-----ELLQERDKLKQOVKKKEEMQOV 1228
QY 1069 CDVAMNIMKSTYLSLESPKDPVLPARFPTQDPKNSNKNYLPPEMKSFETPGKPT 1128
DB 1229 MTTVQN-MQHESAQLOEELHQ---LQAVLVSDSNNSKLQVDY--TGLQSYDQNETK 1282
QY 1129 NVLGAVNRPLSSAQOQFQTS---SRMETVSNASSSSNPSP-----GRIGRLDSS 1177
DB 1283 NFGQELAQVQHSIQCLNTMDLLGLKDIISPOLSSASLLTPQSAECLRASKSEVLSESS 1342
QY 1178 EMDHSENDYTMSSPLDPKKSDDR-----DSDVLVSELEKPPGRKKTPTVDEQEK 1228
DB 1343 ELLQOELBELRKS--LOEKDATTFTQENNHRLSDSIATATSELE---RREHETQSEIK 1396
QY 1229 L---GMDLTQVQEO-----KPKGORSRRRGHTASDQQPEE--KRUKEDI----- 1274
DB 1397 QLKQKQVQLKLEKDKLLIRAKSDQLSSNENNTNKNVNEELLRLQVTLKEKILLEM 1456
QY 1275 ---LENEDQNSPPKGRKRRPPKPLGGGTPEKEPTMKTSSKSKKRSPPAPEE--- 1326
DB 1457 DIGLKGENEKTIVETYGKETE-----YQALQETMKTSMMLREKEPCHSKERKALA 1509
QY 1327 -----EEBERQSGNTQD---KSKSKOHRVSRQAQAE---SPSSAIST--QSTPQK 1373
DB 1510 FEQLLKEKEQGTTELQNLNAVKSQMEKTVVFOQERDQVWALQKQMENTALQONEVOR 1569
QY 1374 GGRPSKTPSPSPQPKN 1390
DB 1570 LRQKFRSNOBELERLN 1586

RESULT 11
US-09-815-242-5834
Sequence 5834, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: EUTRA-011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

```

PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5834
 LENGTH: 2437
 TYPE: PRF
 ORGANISM: Staphylococcus aureus
 US-09-815-242-5834

Query Match 2.6%; Score 185.5; DB 9; Length 2437;
 Best Local Similarity 18.5%; Pred. No. 0.0054;
 Matches 237; Conservative 187; Mismatches 496; Indels 361; Gaps 47;

167 NGHNOKVHMHVDMSSIIICEDTYS-----QELDTVLNVLVP-AHKRL 210
 1433 NGLNQQ---GODLHAKKAINNADVSDVTDIYNNOLDLNDAMETLKHLDNEIPNAEQTV 1488
 211 NKQ-AVLDLAKLL---KRTAQAIIEPYITTFNQVLMGKTSISDSEHVEFDLLELYND 266
 1489 NYQNDNDNAKTNFDDAKRLA-----NTLLNSDNTVNDINGAIOAVNDATIHNL- 1536
 267 SHLLSVLPQLEFKLSNDNERLYQVYKLLAMFGAKOSELASQNKPLMOCTYLGRENDIH 326
 1537 -----NGDQRLQ-----DAKKAQIOSIN---QALANKLKEIE 1565
 327 VPIRECVFASHCLMNHDPDLAKDLLEYLVKVRSHDEEAIIRDVIVSYTA---AKKDI 382
 1566 -----ASNATDQDKLAKKAEPLANSITINIKATSNQESQV 1604
 383 LLVNDHLNFVREERTLDKRRVYRKEMGLAQIYKVALQSAAGKAPAKQIMIKDKLH 442
 1605 QTAGNHAIQVHANEIPK-----AKI-----DANKVDQVQALIDEIR 1644
 443 IYQNSIDRLVERIFPAQYMPHNETTERMKCLYYIATYATDLNAVAKALNEMKQCNLL 502
 1645 NPNLTDKKQALKDRINQILOQGHN-----GINNATKEEIEQAKQL 1687
 503 RHQVLDLIDLIK-----OPKTDASVAKISKVAVITRNLPDQKQODEKKKTQYV- 553
 1688 AQAIDIDKLVAKEDAKODVQVQALIDEIDQ-NPNLTDKK-QALKDRINQILOQGH 1745
 554 -----EDDEKIRKQLEVLVSPSCSQKQEGCVREITKKLGKPNKOPNPLEMIRKL 604
 1746 NDINNALFKKEIEQAKQAQALQALQIDIKDLVAKEDAKNAIKALANAKR----- 1793
 605 LERIAVHIIDTESISALIKQVNSIDGTADDEEGV-PTDQALRAGLELLKVLSTPHIS 663
 1794 -----DQINSNPDLPPEQAKAKLKEIDEAKRALQ-----N 1824
 664 FHSAPFESLACLKMDDEKVAEALQIFKNTGSKIIEEPPIIRSLRLPYLHKKSKGPR 723
 1825 VENAOTIDQLNRGLNL-----GIDDIRNTHVWEVDQP----- 1857
 724 RQAKVAIHCIHAFESKETOFAQIFEP-L-HKSLDPSNLEHLTPVTIGHIALAPDOFA 782
 1858 -----AVNEIFEATPEQILVNGELIYHR-----DITTEQDIAMHMLT--DQIS 1900
 783 APKSWAVAFYVKDILM---DRLPGKTKTKLWVPDEEVSE-TMVKIQALIKMAYRWLL 837
 1901 AEVIDTPSTAFISDLSLAKVEVTLIDGSKVI-VNVPVKVVEKELSVKQQALESJENNAQ 1959
 838 GKKNNHKSQSTLALLTLTILHSDGLDEQGISKPSKRLAAGSAIVKLAQEPCHYE 897
 1960 OKININNSVYTLT-----EQKEAIAAEVYKLL---QQAIDHNDAPDIVS 2002
 898 IITLEOYQCAL-AINDECYQVQVFAQKLIHGLSRLRLPLEYMAICALCARKDPKERRA 956
 2003 VEIIOQOQAHIEQGNPQFTIEQ-----AKSNMISTE 2036
 957 HAROCLVKNINVRREYLYKOHAAVSEKLSLPEYVVPYTIHLLADPDVYVQDIEQLKD 1016
 2037 DAIOHMEIDIKARTDLTDEKQEAATAKINQLEKQAIQ-----AIOQAOSIDE 2083

1017 VKECLMEVLEILMAKNENNSHAFIRKVENIKOTKDAQGPDDAKMKNEKLYTCVYAMNII 1076
 2084 ISEQLBQFAQKKAANPTAKELAKRKQ-EAISRIDFS-----NEKINSTRNSEIGTA 2135
 1077 MRSSTYSLESPPDYPARPFTOPDKNFSNTKNYLLPPEMKSFPTGKRTTNVGLAVNR 1136
 2136 DEKQA-AMNQINEIYLE---TIRDINNAHTLQ---QVEAALNNGIAR---ISAVQI 2181
 1137 PLUSAGKOSQTKSSRRREYTSNASSSSNPPSRIGKGLDSEMDHSENEDYMWSSPLPK 1196
 2182 VYSDRAKOS-----SSYGNESNSHLTIGYGTANHPFNSYTGIRK-----K 2221
 1197 KSDKRDSDLVSRSELEKPRGRKKTPTDEEEKLGMDLLKLVQEQKPKGOSRRRGHTA 1256
 2222 KUDIEDDDIDPLHM-----RHFSNNFGNVYKNAIGVAGISGLASFMTAKRRRK- 2271
 1257 SESDEQWPEEKRLKEDILENEDEONSPEKKRGPRPLGQGTPEKPEPTKTSKGSK 1316
 2272 -EDEBELEIRUNNNDISIKETLDDPKHLPLFVKRRKDEEDDYVEEKDSLNGESLDK 2330
 1317 KKSQP-----PAPEEEEEEQSGNTEOK-SKSKOHRVSRRAQRAESPESALSTOS 1369
 2331 VHTPFFELPKRRRKDEDEVEYTNENTDEKVLKDNHSPILPAKRRKKEED--VEYTT 2388
 1370 TPQKGRGPRSKTPSPQPKN 1390
 2389 IESKDEVDVPLLAKKKKQKDN 2409

RESULT 12
 US-09-815-242-12996
 Sequence 12996, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haseibeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA 011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12996
 LENGTH: 6281
 TYPE: PRF
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12996

Query Match 2.6%; Score 185.5; DB 9; Length 6281;
 Best Local Similarity 18.5%; Pred. No. 0.021;
 Matches 237; Conservative 187; Mismatches 496; Indels 361; Gaps 47;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 20:16:09 ; Search time 25 Seconds

(without alignments)
2354.176 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193
Sequence: 1 MAHSKTRNDGKITTPGCVK.....OKGRGSPKTPSPSPKKNV 1391

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCITUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	232.5	3.2	2662	4	US-09-595-684B-31
2	201.5	2.8	3248	1	US-08-353-700-1
3	201.5	2.8	3248	5	PCT-US95-16216-1
4	194.5	2.7	2482	1	US-08-328-254-6
5	190	2.6	10182	4	US-09-134-001C-3159
6	185	2.6	8991	4	US-08-714-741-32
7	181	2.5	1780	1	US-08-769-309A-5
8	181	2.5	1780	3	US-08-994-570-5
9	178.5	2.5	2368	1	US-08-198-446B-15
10	178.5	2.5	2368	2	US-08-870-693-15
11	176	2.4	1319	2	US-08-290-731C-2
12	176	2.4	1336	3	US-08-290-731C-6
13	174	2.4	1786	3	US-08-973-462-8
14	172.5	2.4	1865	1	US-08-588-985-2
15	172.5	2.4	1865	1	US-08-971-988-2
16	170.5	2.4	3878	4	US-09-914-259-11
17	170	2.4	1187	4	US-08-320-559-28
18	170	2.4	1187	3	US-08-545-860D-28
19	170	2.4	1187	5	PCT-US94-04496-28
20	170	2.4	1210	1	US-08-320-559-26
21	170	2.4	1210	3	US-08-545-860D-26
22	170	2.4	1210	5	PCT-US94-04496-26
23	168.5	2.3	3433	4	US-09-091-501B-10
24	168	2.3	564	3	US-09-308-022-6
25	166	2.3	947	4	US-09-418-780A-1
26	166	2.3	1231	4	US-08-714-741-41
27	165	2.3	1333	3	US-09-356-952-2

28	165	2.3	1850	4	US-09-620-093A-5	Sequence 5, Appl1
29	164.5	2.3	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
30	160	2.2	1312	4	US-09-345-882-29	Sequence 29, Appl
31	158.5	2.2	714	2	US-08-990-114-3	Sequence 3, Appl1
32	158.5	2.2	714	4	US-09-241-333-3	Sequence 3, Appl1
33	158.5	2.2	1507	4	US-09-914-259-37	Sequence 37, Appl
34	158	2.2	2474	4	US-08-265-967C-3	Sequence 3, Appl1
35	158	2.2	2474	4	US-08-305-790B-4	Sequence 4, Appl1
36	158	2.2	2972	4	US-09-579-181-2	Sequence 2, Appl1
37	158	2.2	3118	4	US-09-579-181-1	Sequence 1, Appl1
38	157.5	2.2	1111	4	US-09-914-259-28	Sequence 28, Appl
39	156.5	2.2	414	1	US-07-667-276A-4	Sequence 4, Appl1
40	156.5	2.2	1588	5	PCT-US93-07261-11	Sequence 11, Appl
41	156.5	2.2	1663	5	PCT-US93-07261-16	Sequence 16, Appl
42	154.5	2.1	631	3	US-08-847-065-25	Sequence 25, Appl
43	154.5	2.1	1848	3	US-08-296-791-6	Sequence 6, Appl1
44	154.5	2.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl1
45	154.5	2.1	2101	1	US-08-466-390-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-595-684B-31

Sequence 31, Application US/09595684B

Patent No. 6544766

GENERAL INFORMATION:

Query Match

Best Local Similarity 3.2%; Score 232.5; DB 4; Length 2662;

Matches 274; Conservative 270; Mismatches 540; Indels 425; Gaps 64;

QY	17	PGVKISIKISKEEVRRLKMYVKTFFMDODDSE-----EKELYNLAHLASD	66
DB	333	PYVNEVS---IDELALKRR---KEIMDKKLEVSLETRAQAMEKD---QLADLLEK	383
QY	67	FLFKHGKDVRLVLAACLDIFRIYAPAPYSPDKLIDIF-MFTROLKGL-EDTKSPQ	124
DB	384	DLLQK-----VQNEKIENLFFRLVLTSSSLFLOGLKAKR	417
QY	125	ENRYYLLENIAWAKSYNICELEDSNELFPOLYLTLSVINGNNOVHHMVDLMSSI	184
DB	418	KRRVYWCGLGKINKMNSN-----YADQFNIPFN-ITTTTHKLSIMLREI	461
QY	185	---ICEGDTVSGELDPVL-VNLVPAHKNLKNQAVDLAKALKRQAQNIIEPYITFFNOV	240
DB	462	DSVCSSEDSVSNITDITSEIEMNPATILNDENIE-----	497
QY	241	LMKTSISDSIEHFVLDILELYNIDSHLLSVLPOLFEKLSKSNDEERLQVKKLAKMF	300
DB	498	-----SELNSLRADVDNLVL-----DYQLRPREKEMLKLEKKNDDLEFALERKTR--	545

```

QY 301 GAKSELASONKPLMOCYLGFRNDIHPVIRLECVKASHCMLMHPDLAKDITEXIKVRSR 360
D 546 --KDOEQ-----LIHISNKNKNVYKREYVNO 571
QY 361 DPEAIRHDIYIVTAAKDDILLVNDHLNFBVRERTLDKRWRRKEMMGLADLYKKA 420
D 572 DLEBEL-----SKVELREKEQOLKLOEYDSOK--LENTIKMOLS-----YS 613
QY 421 LOSAGKDAKQIAMIKDKLHIIYONSIDRLVERIFAQYVWPHNLEETERRMKCLYLL 480
D 614 LESIEDPKOMQOTLEDAETV-----ALDARESAPFLSENELKEKKELATAT 661
QY 481 YATIDLAIVKALNEMKQONLIRHOVDLDTLOPKTDASVKAIFSKVMYITR----- 534
D 662 YKOME-----NDIOYQSOLEAKKMQVDLEKE-----LOSAPFNETIKLTSIDGKV 708
QY 535 -----NLDPGRKQODEMKKFTOYLEDDEKIRKOLEVLYSPTCGKQAGSCVREITKLG 588
D 709 PKDILCNLEBEGKTIDQKELNKVEEVEALREEV-ILLS--ELKSLPSEVERLRREI- 763
QY 589 NPKOPTNPLEMIKFLERLAPVHIDTES-ISALIKOVNKSIDGTADEDEGVPTDOAI- 646
D 764 ---QDKSEELHIITSEKDKLEFSEVYHKESRYQGLLEBEGTKKDDLATTOGSYKSTDQEFQ 820
QY 647 -----RAGLELIKVSFTHPISFHSATFESILLACKMD----- 680
D 821 NFKTLHNDPEQYKMYLEENERRMNOEIVNLSK-----EAKFDSISLAKTLELSYKT 872
QY 681 ---DEKVAEALQJFKMTGSKIEEDPPIHRSALLPVLHKS-----K 719
D 873 QELOEKREVOERL--NEMEOUKOLERDSPLQTVBEREKULITEKLOOTLEEVKTLTOE 930
QY 720 KGPPOAKYAIHCHIAIFSSKETQFAQIFEBLHKSIDPSNLEHTITPLVTIGHTALLAPD 779
D 931 KDKDKLOESLQI-----ERDQKSDIHDVTVMNIDPQ--EQLRNLESLSKH-----Q 976
QY 780 QFAPKMSVATFVYKDLMDRLPGKTTKLVMPDEVESEVETVAKLOAIMMYRWLLGM 839
D 977 ETTITLKSISEEVSRLHM-----BENTGETDEFO-----QKMWGI 1014
QY 840 KNNHSGSTGLRLTLILHSDGLTEQGISK--PMSRLRLAAGSAIVKLAQ-- 891
D 1015 DKKODLEAKNT-QTLTADVKNETIIEOQRKTFSLIOEKNELOQMLSEYIAKEBKLKDDK 1073
QY 892 EPCYHEITTEQYQCALAINDECYQYVQFQRLHKLGRRLRPLEYMAICALCA--KD 949
D 1074 ENIEMTIENQBELRL-----LQDELKQOEIYAQEKNAIKR--EGELSPFCDBLAEE 1126
QY 950 PVKERRAHARQCLVKNINVRREYIKOHAAYSE-----KLSLLEEYVVPYTIHLLAH 1001
D 1127 KLEKSSOQLOEKQOOLINVOEEMSEMOKKTINEIKNELKNELTLEHMETERLELAQK 1186
QY 1002 -DPPYVAVODI-EOLKDVKECJLWFLVETILMAKNENSHAFIRKAVENIKOTKDAQGPDA 1059
D 1187 LANETVEEVKSTIKRKYLKE-----LOKSPETERDHLRGYIRELEATGLQTKELKTAHI 1241
QY 1060 KMNELTYVCDVANNIIMSSTYVLSLSPKDPVLPARFFTOPDKNFSTNKVYLPPEKSF 1119
D 1242 HLKEHOETIDELRRSY--SEKT-----AQIIMQDLEKSHTK--LQEEIP-- 1282
QY 1120 FTGQKPTTVNLGAVNKPSSAGKQOTKS--SMMEVYVSNASSNSSPKRLINGRDDS 1177
D 1283 -----VLHEOEELLPVYKVSFQETMNELELTITRSTYTKDSTYIARI----- 1325
QY 1178 EMDHSE-NEDYTMSPLPCKKSDKSDSDVLRSELEKPRGRKKTPEV-----Q 1225
D 1326 EMELERLNKEKFOESQEIKSTIKRDMUKITIKALEVKHDKLHIIETLAKIOESOSKQ 1385
QY 1226 EELKGM---DILTKLIVQEO--KPK-----GSGORSKRRGH-----TASBSE 1261
D 1386 EOSLNMKEKNETKIVSEMOQFPRKDSALIRIEMIGLSKRLQESHDEMKVAKKEDD 1445
QY 1262 QQ-----WPEBKRLKEDI-----LENEDEONSPFKKGRGKRP--KPLGGGTPKEEPT 1307

```

```

D 1446 LQRLQEVLOSDDQKNIKIEYAKHLETEBELKVAAHCCLEQDETINELRVNLSEKETE 1505
QY 1308 MKTSK-----GSKKSGPPAPEEEDP---EROSGNTQEKSKSGKQVRSRQAORA-ES 1358
D 1506 ISTQKOLEAINDKLQNIQEIYERKEBOLNIKOISEVQENVNELKQREHKKAKDSALQS 1565
QY 1359 PESSAIEST 1367
D 1566 IESKMELEL 1574

RESULT 2
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 2.88; Score 201.5; DB 1; Length 3248;
Best Local Similarity 19.5%; Pred. No. 1.2e-06;
Matches 298; Conservative 234; Mismatches 590; Indels 407; Gaps 63;

QY 19 VKESIDKI-----SKEMVRLKMWVTFMDMODSSEKELY----- 56
D 1900 MKELDSKTLHJQEVQMLKIEACIELEKIVGELKKNESDLSKLEYFSCDHQELLQRYETS 1959
QY 57 -LNLALHLASDFLKHGKQDVRLVACCLADIFRIYAPAPYTPSPDKLDIFMFTIRQL 114
D 1960 EGLNSIDLEMHAD--KSSREDIGDNVA-----KVNDSWKERFLDYENEL 2000
QY 115 KGLDETKSPQFNRYFYLLENIAMWKSYNICFELSDSN--ELFQLYTTLPSVINNGHNOK 172
D 2001 SRINSEKASIEHEALYLEADLEVVOTIEKLCLEKDNENKQKVIYCLEBEL--SVVTSERNOL 2059

```

```

QY 173 VHMHWDLMSSTICBGPVSOELDTVLNVLPAHKNLNKQAVDLAKALLKRTAALIEPY 232
DB 2060 --RGELDTWKSRTTLLDOLSEKMER--TQELSHQSECLHCIOVAEAVKERTELLQ-- 2113
QY 233 ITTFNOVLMKTSISDLSEHVPDL-----LELYNIDSHLLSLVLPOLFEKLXSNDE 287
DB 2114 -----TSSDOVELLKDKTHQOEKLOSLKENDSQALSILTKCELENOIOLNKE 2160
QY 288 ERLQV--VKLLAKMFAGANDELSAONKPLMOCYL--GRENDIHVPRLIECVFASHCLMN 343
DB 2161 KELLAKESESIOLARSESDYEKLAVNSKALEALVEKGEF-----ALRLSTQEEVHQQLR 2215
QY 344 HPDLAKDLLEYKLVSHDPEEARLHDVYSITYTAKKOLLVNDHLNVRERTIDKR-- 401
DB 2216 G-----TEKLRVIEADEKQOLH--IAEKLERERENDSLKDVENIERELOXSEENO 2266
QY 402 -----WVRKEMMGLAQIYKRYALQSAAGKD--AAKOJAWTKDLIHTY 444
DB 2267 ELVJIDAKNSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIOEKQGLSELD 2326
QY 445 YONSIDDLILVERIAQIYVWPHNLETTERMKCLYLYATLDLNAVAKALNEMKCONLRLH 504
DB 2337 KLLSFSKSLKEKEQAEIOIKESKTADEM-----LON 2359
QY 505 QVKDLIDLIKOPKTDASVKAFISKVMVITRNLDPD-----GKAQDFMKFTQVLDEDEKI 559
DB 2360 QKELNEAVNALCGD-----QELMKATEOSLDPIIEEHOLRSTIEKLRLARLEDEK- 2411
QY 560 KROLEVLVSPSCSKQAEQVREITKLGPNKOPNPFLMIKFLERITAPRHIDESTIS 619
DB 2412 -KOLCVLQ--QKLESEHMDLKGVRNLERE-----LEIRTNQENHAA--LEAENK 2459
QY 620 ALIKOVNSIDGTADDEBQVPTDOAIRAGLELLKVLSTFHPDISHSAFESSLACLKM 679
DB 2460 GEVEITLAKIRIGMT-----OSLR-GLEL-----DVYITRSEK--ENLNLNELOK 2499
QY 680 DDEKVAEAL--QIEKNTGSKIEEDFPHIR--SALLPVLHHSKSKPPROAKVIAHCH 734
DB 2500 EOEKISELEITINSSSENILQOEKQVOKKESSTAMELOLQULELNERVA--ALHNDQ 2557
QY 735 AIFSSKETQFOAIEPRLH-----KSLDPSNLEHLITPLVYIGHIALAPDOFAAPWKS 787
DB 2558 EACKKKEGNLSQVCLLEKRAQLOGLDEAKNNIYVLOSSKG----- 2601
QY 788 WVAFTIYVDLMDNRLPGKKTITKLVNPDEEVSPEPMVKIQAIKMVRMLGKNHNS--K 845
DB 2602 -----LIQVEDEGQ--KLEKKDEEIS-RLKNQIODEQVLVSKLSQVEGEHQLMK 2648
QY 846 SGTSLRLTLTLHSDGLTEQKISKPDMSRLRLAAGSAIYKLAQEPCHHITLLEQY 905
DB 2649 EONELRLNLTVELEOKIOVLOSKNASLD-----TLEVLQ 2683
QY 906 LCALAINDECYQV--ROVFAOKLHKGLSR--LRPLVEYMAICALCADDY--KERRA 956
DB 2684 SSYKLENELELTTKDKMSFYKAKMTAKETELDREHMEMAKTALDEBELSGEKNRIA 2743
QY 957 HAROCLVKNINVRREYKLOHAHVSEKLSLAP-----EYVVPYTHILAND 1002
DB 2744 GELQILLBEIKSSKQOLKELTLENSELKSLDCMHKDYERKGVREELAEVQLRL--HE 2801
QY 1003 PD-----YKQVODI--EQLDVKECL--WFVEIIMAKNE--NNSHAFIRK 1042
DB 2802 AEKKHQALLDLTNKOYEVEIQTYREKLSKEECLSSOKLELDLLKSSKEELNNSKATQ 2861
QY 1043 VVENIKQTKDAQPPDAKMNKELTYVCVA--ANIIMSKSTYSLESPKDPV-----L 1093
DB 2862 ILEELKTKR---MONLKYVNOJLKKENERRAQCKMKLLI--KSCKOLEEKEKLIQDELSQL 2915
QY 1094 PARFTOPDKNFSNTR--NLPPEMKSFTTPGKPKTTNV-----IGAVNKPILSSACK 1143
DB 2916 QAAQEKQKTGVMOTKVDLFTTEIKELKETEKEKTEADEVLDKYCSLLISHEKLEKAKE 2975

```

```

QY 1144 OSOTK-----SSRMETVSNASSSSNPSPGRIGKRLDSMDHSENEEDYTMSPLP----- 1194
DB 2976 METQVAHLCSQSQSKODSNGSLPLGVPVG-----PSPITPSYE 3014
QY 1195 -----CKSDKRDSDLVRSLELEKPRGRKTPYTEDEKLGMDLTKLVQEQRPKGSQ 1247
DB 3015 KRLSSGQNNKASGRQ-----RSSGIWENGRGPTPAT-----PESFSKSK 3054
QY 1248 RSRKRG-HTASDEQOPEEKRLAKEDIENEDONSPPKKGKRGPRPKPLGGTIP--KE 1304
DB 3055 KAVMSGIHPA-----EDTEGTEFEPEGLPYEVVYKGFADIPGTGTSPTYLIR 3099
QY 1305 EPTMKT-----SKKSKKSGPPAPDEEPEERQSGNTQESKSKQHRVSRRAQRAESPSS 1361
DB 3100 RTTMAITRPSRLAQAOKLALSPSLGKENIAESSKPTAGSRSQKVAQOR-----SPVD 3153
QY 1362 SAI-----ESTQSTPOKGRGRPSKTPSPSQ 1386
DB 3154 SGTILREPTTKSVPNNLPERSPDPSPRE 3182

```

```

RESULT 3
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ANTI-SENSE: NO
; HYPOTHEICAL: NO
; PCT-US95-16216-1
;
; Query Match
; Best Local Similarity 19.5%; Pred. No. 1.2e-06;
; Matches 298; Conservative 234; Mismatches 590; Indels 407; Gaps 63;
;
; QY 19 VKESIDKI-----SKEENVRRLKMWKTPMDMDQDSEEEKELY----- 56
; DB 1900 KNEIDSKLHDEVOQJMTKIEACIELEKTIYVGLKKNESDSEKLETFSCDHQBLRQVETS 1959

```


Best Local Similarity 19.4%; Pred. No. 2.8e-06;
Matches 297; Conservative 233; Mismatches 592; Indels 407; Gaps 63;

```

OY 19 VKELSDKI-----SKEENVRLKMYVTFMDMDODSEEEKELY-----56
Db 1172 MKELDSKHLQEVOLMFKIEACIELEKIELEKKNENSDLEKLEFESCDHOELLORVETS 12331
OY 57 --LNLALHLASDFLHPGKDVRLVACCLADIFRITVPEAAYTPDKLQDFMFTIRQL 114
Db 1232 EGLNSDLEKHDAD---KSSREDGDNVA-----KYVDSWKERFELVENEEL 1272
OY 115 KGLEDTKSPQFNRYFLLLENIAWVSYNICELEDSN--EIFTQLYRTLESVYINGNHCK 172
Db 1273 SRIRSEKASIEHEALYLEADLEVOTFEKLEKKNENKQVYVCELEEL-SVVTSEKRL 1331
OY 173 VHMHVADLMSSITCGDTVSOELDTVLVNLVPAHKNLKNQAYDLAKLLKRTQAQAIPEY 232
Db 1332 --KGLDTPMSKTTALDQLSEKMEK--TOELSHQSECLHCIOVAEAKEVEKETELLO-- 1385
OY 233 ITTFNOVLMIGKTSISDLSHEVFDLI-----LELYNIDSHLLSVLPQLEFKLSNDNE 287
Db 1386 -----TSSDVSELLKDKTHLOEKQSLSEKSOALSITKCELEENQALQNKKE 1432
OY 288 ERLAQ--VKILAKMFGAKDSELASQKPLMOCYL--GRFNDIHVPIRELCVYFASHCLMN 343
Db 1433 KELVYKESSEIQARLSESDYEKLNVSKALEALVYKGF-----ALRISTOEVEHQLR 1487
OY 344 HPDLAKDILEYKVRSHDPEEARHDIIVSYVTAAKDILVNDHLNLFVBERLTDK---400
Db 1488 G-----IEKRLVYIEADEKKQLH--IAEKLERERENDSLKDKVENLRLQSEBNQ 1538
OY 401 -----RMVVRKEAMGGLAQIYKKYALO-----SAAGKPAKQIAMIKKLHITY 444
Db 1539 ELVILDAENSAEVEETLQIIEMARSLKVEELDVLTRSEKENLTQIOEKQGLSELD 1598
OY 445 YONSIDRLVLERIFQAYVPHNLLETTERMKCLYLYATLNLNAKALNEMKCONLRLH 504
Db 1599 KLLSEFKSLLEKEBQAEIQIKEESTVEM-----LQN 1631
OY 505 QVKDLDLILKOPKTDASVKAIFSKVAVITRNLPRD---GRAQFMKKTVOLEDEDKI 559
Db 1632 QLKELNEVVAALCGP-----QETMKATEQSLDPIEEHQRLRSIKELARLEADER- 1683
OY 560 RKOLEVLVSPTCSCQOAGCVREITKKGKQRPINPFLMKILLENIAVYHIDTESIS 619
Db 1684 -KOLCVLO---QKSESHHADLKGRENLERE---LETARTQOEIAA---LEAENSK 1731
OY 620 ALIKOVNKSIDGTADDEDEGVPTQAIKAGLELLKVLSTFTPIFSHSAFESLACIKM 679
Db 1732 GEVELLAKKIEGMI-----QSLR-GIEL-----DVTIRSEK--ENLTNELOK 1771
OY 680 DDEKVAEAL--QIFKNTGSKIEEDFPIR---SALLFVLNHHKSKGPPROAKYAHCIH 734
Db 1772 EOERISELEIINSFENILOKEQEKVQMKESSTAMEMLOQKELNERVA--ALHNDQ 1829
OY 735 AIFSKKEQFAQITEPLH-----KSLDPSNLEHLITPLVTIGHIALALADQFAAPKRS 787
Db 1830 EACKAKQONLSSQVCELEKQAOLLOGLDEAKNNYIVQSSVNG-----1873
OY 788 WVAFTVYKDLMLNRLPGKTTKTLMPDEEVSPTMYVQIKMMVRLMLKMNHNS--K 845
Db 1874 -----LIOEVESQK--KLEKNEBIS-RKKNQIOQOQOLVSKLSQVEGEHQJMK 1920
OY 846 SGTSTLRLLTTLHSDDGLTEQGIKSKFMSRLRLAAGSAIVKLAQEPCHIEITLLEQY 905
Db 1921 EONLELRLNLFVLEQKIOVLQSKNNSLOD-----TLEVIQ 1955
OY 906 LCALLINDECYQV---ROVFAOKLHKGLSR---LRLEPVAMICALCAKADPV---KEBRA 956
Db 1956 SSYKMLELELTLMDKMSFVEKYVAKMTAKETELQREHNEHAQYTAELQOEELSGEKNLA 2015
OY 957 HAROCLVKNINVRREYLKQHAAVSEKLSLTP-----EYVVYTTIHLALND 1002

```

```

Db 2016 GELQLLLEBEIKSSKDQLKELTLENSELKSLDCKMHKQDVEKGVREIEAYQLRL--HE 2073
OY 1003 PD-----YKQVDI--EOLKQVDECL---WFLVEITLMAKNE--NNSHARTK 1042
Db 2074 AEKHOALLDPTNKQVEYEIQTIRKILTSKECLSSQKLEIDILKSSKEELNNSLKATQ 2133
OY 1043 MVENIKOTKDAQGPPDAMKNEKLYTCOVA---NNIIMSKSTVYLSLESPDPV-----L 1093
Db 2134 ILEELKTKR---MDNLKYVNLQKKNERAGAKKLLI--KSCQOLEBEKEILLQKELSOL 2187
OY 1094 PAREFTQPDKNFSNTR--NYLPPEMKSPFTPKPPTTVN-----LGAVNKPLSSAGK 1143
Db 2188 QAAEQKQKGTVMQTKVDELITTEIKLEKETLEKTKAEDEVLDKYCSLLISHKELEKAKE 2247
OY 1144 OSQTK-----SSRRETVSNASSSNPSSPGITKGRILDSSEMDHSENEDEYTMSSLP- 1194
Db 2248 MLETOVVAHLCSQOSKOSRGSPLGLPVVPG-----PSDIPSVTE 2286
OY 1195 -----GKSKDKHDDSDLVSELEKPRGRKKTPTVBOEKLGMDDLTKLVQEQKPKGSQ 1247
Db 2287 KRLSSGONKASKGKQ-----RSSGIWENGSGPTAT-----PESFSKSK 2326
OY 1248 RSRRKG--HTASESEDOQWPEEKRLKEDILENEDQNSPPKKGRGRPPKPLGGCTP--KE 1304
Db 2327 KAVMSGIHPA-----EDTEGTEPEPEGLEPVYVKKGFADIPGKTSPTILR 2371
OY 1305 EPTMKT---SKKSKKSGPPAPEEEEEEROSGTEOKSKSKOHRVSRRAQRAESPES 1361
Db 2372 RTVMATRTSPRLAOKLALPSLSTGKENLAESSKPTGAGSRSSOKVYKVAOR-----SPVD 2425
OY 1362 SAI-----ESTQSTQKGRGRSKTPSPSQ 1386
Db 2426 SGTILREPTTKSVPVNLPERSPTDSPE 2454

```

```

RESULT 5
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelle-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

Query Match 2.6%; Score 190; DB 4; Length 10182;
Best Local Similarity 19.1%; Pred. No. 6.8e-05;
Matches 276; Conservative 235; Mismatches 523; Indels 414; Gaps 67;

```

OY 23 SDKISKEEMVRLKMYVTFMDMDODSEEEKELYLNALHLASDFLKHGKDVRLVAC 82
Db 8944 NDEITMNTNSIAQLSRYVQAQFLLDANKSLDELNNQAFVQAASSNYINSDCLKQOQDI 9003
OY 83 CLADIFRIYAPEARYTSPDKLQDFMFTIRQLKGL--EDTKSPQFNRYFLLLENIAWV 138
Db 9004 ALSNARVYLAKENKNKNDK-----QIOGLKQVIEDTKO-----9037
OY 139 KSYVIFCELEDSNEIFTQLYRTLESVYING--HNQYHMHMVDLMSSI-----ICEGDT 190
Db 9038 -ALNGIORLSKAKAKAIQYVOST--SYINDAQRIHAENNINHSIDLSSIAVTLTKASADLDN 9095

```

```

QY 191 VSOELDTV-----LVNLVPAHKNLN---KQAYDLAKALLKRT----- 225
Db 9096 AMKDLBRTIESNSTSVPNNSVNYINADKNQIEFDEALQOASATSSKSENPATIEEVLGL 9155
QY 226 AQAII-----EPITTFNOVLMAGTISIDSEHVEDLLELYNIDSHLLSVLPQ 276
Db 9156 SQAIYDTKNALNCEORLATEKSKDLKLIK-GLKDLNKAQOLEDVYN--KVSANTLLELSQ 9212
QY 277 L-EFKLKSNDNEERLO-VVKLLAKMEGAK-----SELASQ--NKPLMOC--YLGRENDI 325
Db 9213 LTOSTLELNDKMKLTKLTKTLVNPVKASLNYRNADYNLKRQFNKALKEKGVYLNKNGSGT 9272
QY 326 HVEIRLECVAFASHCLMNHPLAKD--LTEXLYKVRSHPEBAI-----RHDIYSIVTA--- 377
Db 9273 NVNII-----NDIOH--LITQIDNAKQDQNGERRLKEHQKSEVFLIKLELDLNNNAQKAAII 9326
QY 378 -----AKKDILLVDHLNFEYRETLDKRMVREKEMGLQITKTAALQ--SAAGKDAKQ 432
Db 9327 NOIRASDKIKIINOYDNAME-----LNDAMQGL-----KEHYAQLTATTKD---- 9368
QY 433 IAWIKDKLHIYONSIDDLLEVERLFAOYVPHNLE-----TEEMKCLYLYAT 483
Db 9369 -----NIEYLNADDEHKL-----QYDYAINLANVLDKENGTKNKANIIIGNIQN 9413
QY 484 LD-----LNAVKALNE--MMKQNLRRQVRLDLIKQPTDASVKAIFSVVYITERNLP 537
Db 9414 MDDARALLNGIERLKAQOTKAHNDIKDTLRQJDEIHNAMATSKSA--QAKQWY----- 9466
QY 538 DPGKADFMKFTQVLEDEDEKIRKOLEVLYSPICS--CKQA--EG-----CYREITTK 586
Db 9467 -----NEARKALSNINDATSNLDVNAKQEGOSAIEHINADELPA 9508
QY 587 LGNPQPTNPFELEKIKELERIAVHIIDTESISALIKQVKSIDGTADDEGVPTQDAI 646
Db 9509 KIDANQMDQKVEDINHLISQ--NPNLSNEKKKLISQIKLVNGINELIQOAIKQOIE 9566
QY 647 RAGLELLKLVSTFHPISTFSAETFEESLACLAKMDEKVAEALQIEKNTGSKIEEDPHI 706
Db 9567 NATTKDEVIETETTKLIIAKAE-----AKOMIKELSQKKRD----- 9602
QY 707 RSLALPLVHLHKSKKPPROAKVAIHCHIAIFSSKETQFAOIFELHNSLDP--SNLEHII 764
Db 9603 -----AINNNTDLTPSOKA-----HALADIDKTEKDALOHEN--SNSIDIDINNKCHAF 9650
QY 765 TPLVITIGHIALPADQ-----FAAP-----MKSVAATFIV-----KDLIMDRL 803
Db 9651 N---TLAHIIWPTDQPLVEFVPELSQNALVITSEVVARDETISTESIKKKTITLDEL 9707
QY 804 PGKKTTLKWPD--EVSSEPTMKIOAI-----KMAVRWILGKNNHKSQSGTS 849
Db 9708 ---KVNIVSLPNTDKVDHDLTAKVYKVLADGSYTVVNPVKVVEKEIQIAKKDAIKTIDV 9764
QY 850 TLRLTLTIHSDDDLE--QGIKISKRPMSRLRLAAGSAIVYLAQEPCHETIT----- 900
Db 9765 LVROKIKIDIDSNNEELISTOREDAKAELEK--KQALIDVNNHKSIKIDLEYVAKRTDFEE 9821
QY 901 LEOYOLCALAIN-----DECYQVROVPAQ--KLHGLSRLRLPYEYMAICALCANDPV 951
Db 9822 IDGFDPKRFITLNAKAKKIITDVMTQIONGKEIETIKGLTSNEMTQDQKOLTAQKFELE 9881
QY 952 KERRAARQCLVKNINVRREY-----LKQHAASVSKLSLLEPYVVPYTHILLANDPD 1004
Db 9882 KVEHAHN---LVEINQOQEEFNRRYKHILNQAHILGK-----HIAEKKLG 9924
QY 1005 YVAVQDIEQKDYKECLMPLYE-----ILMAKNENSHAFIRKAVENIK 1048
Db 9925 YVAVNKTQOILINNOASAFYIKQWALDRIKOILETMNSIRGAHTVQOVHALLQIGIOIL 9984
QY 1049 QTKDA-----QGPDAKMEKILTYCDVAMNITISKSTYTSLESPDPO 1092
Db 9985 KVNASTINOSFNDSLHNFYLSKFDARLEK-----DVANHIVOTETFEVYLGK--TGV 10037
QY 1093 LPARF--FTQPDKNFSNTKNVLPPEMKSFF--TPGKPKTNTNVLGAVNKPLSSACKOSQTK 1148

```

```

Db 10038 EPCKINKETQOQPLKHNNDSDJFKHLVDNFGKTVGVITLTGLISSFWLAKRRKEE 10097
QY 1149 SSRMEVSVNASSSSNFGRIKGRIDSSSEMDHSEMEDYTMSSPLPGKSSDKRDSDDLVR 1208
Db 10098 KQSK-----NHHKDIRLSD-----TDKIDPIVITRK 10123
QY 1209 SELEKPRGRKRPVTEQEKIGMDLTKL--VOEQKPKGSQSRKRGHT-----ASESD 1260
Db 10124 RKIDK-----EEQIQNDKHSIPVAKKKSKEKQJSEEDIHSPVKKRQNSD 10171
QY 1261 EQQPEEK 1268
Db 10172 NQRYKTEE 10179

```

```

RESULT 6
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
;
GENERAL INFORMATION:
; APPLICANT: Biles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swatlo, Edwin
; APPLICANT: Yochev, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-0712
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8991 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
;
US-08-714-741-32

```

```

Query Match 2.68; Score 185; DB 4; Length 8991;
Best Local Similarity 20.08; Pred. No. 0.00014;
Matches 197; Conservative 120; Mismatches 322; Indels 348; Gaps 46;
QY 538 DP-CKADFMKFTQVLEDEDEKIRKOLEVLYSPICSKQAKGCVREITTKLGPKO--PTN 595
Db 8210 DPGKTDDELDKRAA-----BAELNKVVEALLPNQVALEIE-----ELSKLELDNLKQAEITN 8259

```


Sequence 5, Application US/08994570
Patent No. 6090929
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauack, Theresa M.,
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,570
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090929and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SRO ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5

Query Match 2.5%; Score 181; DB 3; Length 1780;
Best Local Similarity 20.9%; Pred. No. 2.1e-05;
Matches 99; Conservative 59; Mismatches 151; Indels 164; Gaps 19;

QY 1036 SHAFIRKMWENIKQTFKDAQGDADAKMNEKILTVCDVAMNIMSTYSLSPDPVIPA 1095
DB 247 SHAETSPAESGQAVECKEKEEK-----QEKPSASPSPTSPVT-- 288
QY 1096 RFTQPDKNFSNTKNYLPPKMSFFTPG-----KKTNTVLGAANKPLSAGKQ 1144
DB 289 -----SETGS---TFKKFTQGMAGMRKTSFRKKEDEVEASEKKKQDEPEKV 334
QY 1145 SQTSSRMETYS---MASSSSNPSSPG-----RIKGRDSSSEMDHSEDEDYMS---- 1190
DB 335 DTEEDGAEVAASEKTLTASEQAHPESAHEPRLSAEYKVELPSEBOVSOGSPSEK 394
QY 1191 -SPLPGKSKDKRD-----SDLVRSLEKPRGRKKTPTVE-----QEEKLMD----- 1232
DB 395 PAPLATEYFDEKIEVHOEYVAEYHVSTVEERTEDOKTEVEETLGSVPAAELVGDAPQ 454
QY 1233 -----DLTKLVQ-----EOKPKGSORSRRKRGHTASESDEQMPPEKRLKE----- 1272
DB 455 EAEPKALVKLKEFCVSGEDPTQCADLS-----PDEKLSKPPPGVYSEV 499
QY 1273 DILENEDE---QNSPPK-----KGKGRPPKPLAGGT--PKEPTMKTS 1311
DB 500 EMLSSQERMKVQSGPLKLTFTSTGLKLSGKKGKRGSGDESGEHTQVPADSPDSQEE 559
QY 1312 KKGSKKSGPPAPE-----EPEEROSGNTQKSKSKO----- 1345
DB 560 OKGSSASSPPEPEITCLEKGLAEVQODGAEGGATSDGKKREGVTPWASFKKMTPTK 619
QY 1346 HRVSRRAQ-----QRAESPESAISTOSTPOKGRGRPSKTPSPSQPKNV 1391

DB 620 KRVRRPSESDKEDBLDKVKSATLSTESTASQJEMKGSVEBPKEPKEKRV 672

RESULT 9
US-08-198-446B-15
Sequence 15, Application US/08198446B
Patent No. 5674996
GENERAL INFORMATION:
APPLICANT: Hartwell, Leland H.
APPLICANT: Weiner, Ted A.
APPLICANT: Pilon, Sharon E.
APPLICANT: Groudine, Mark T.
TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/198,446B
FILING DATE: 18-FEB-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Shelness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCRI7537
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-224-0779
INFORMATION FOR SRO ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2368 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: yeast MEC1 protein
US-08-198-446B-15

Query Match 2.5%; Score 178.5; DB 1; Length 2368;
Best Local Similarity 19.4%; Pred. No. 5.5e-05;
Matches 251; Conservative 188; Mismatches 477; Indels 377; Gaps 60;

QY 68 FLKHFGKD--VRLLVACCLADIFRIYAPPAEAPYSPDKLQIFMFTITQLKGLGDETKSPQF 125
DB 526 FLSHNPIDEPSESLISGLTFLSLHRTFS---HQPPRLTGGNGQINSFV-LVQKCFMNS 560
QY 126 NRYFYLLFNIAWVSYNI--CFELEDSENEIFTOYRPLFVYINNGHQAQKMHMVDLMSS 183
DB 581 NRYIRLL-----STIIIPLFNISDSHNSDEHTALILFLQSQKIPVYKENVIAWTQ 633
QY 184 IICGDTVSOELDTVLNVLPAAKN-----LNKQAYDLAKALKLRTQAQALEPYITT 235
DB 634 LTL---TTSNDVFTLLKLILDFNSDDYSLRIWMTLQIRNMAKILKKTPTVQLSLPILPV 690
QY 236 FFNQVLMGKTSIDSEHVFDLIELYINDSHILLS-----VLPQLEFKLSNDNFERL 290
DB 691 LLRQ---LGRNLVE--RKGEGQNLIELGYPSKTIIDIFQRYIIPVAIIDYKSD----- 739
QY 291 QVVKLAKMFGAKSELSAONKP-----LWQCYLGR---F 322
DB 740 -VLSEIRKINCDGFTSLINQKQVLLKKNRQIFAVALVKGHGLFSDILLETPLNRAPTF 798
QY 323 NDIHVPIRLBCVK-FASHCLMNHDLAKD-----LTEXLKVRSN----- 360

```

Db 799 DKGITAVLPDYKTLAEITKLYKNSVTKDASDENANMILCSLRFITNFEKDKRHSKY 858
QY 361 -----DPEAIR---HDVIVSVTAAKRIDLLYNHDLNPFVEERTIDK-----RRRV 404
Db 859 KNINMTDQEQAFKKLQDNITLGIEQVFSSDIHVEGRTTYEKLRYINGISFLIYAP 918
QY 405 REAMMGIAQIYKKYKALOSAAAGDAKAWIKDKLHLYONSIDRL--IVERIFA-- 460
Db 919 KKSIIISALAOI--SICLOTGLG---LKEVRYEAFKCMHLVHNLDEELSTYIDSLAFI 973
QY 461 -----QYVVPNHEETTERMKCLYTL-----YATLDLNAVALNEMMKQNLRL-- 503
Db 974 LQKMFSENGKLRNIYVSTIDPLIKREKSDILKRPYTLALVGRKELGILANDGOFARVY 1033
QY 504 HOVKRLDLIKOPKIDASYKAIFSKVAVITRNLP-----PGRAQ 543
Db 1034 NKIRSTDPLI--PIFANNIKS--SNKYVINQDLQIEVYLRRKQTERSIDFTPKKVGQTS 1089
QY 544 DFMKFTQVLEDEDEKIRKQLEVYVSPSCQKQAEQVREITPK-----LGNPKQPTNP 596
Db 1090 DITLVGLALDLSHFRNLDKDLCKCAKICSMIG-VLDYTKHEKRRTTYSENEVYDLND 1148
QY 597 FLEMIFLTL---ERIAVPHIDTESIS-----ALIKQVNSIDGTADD---EDEGVPT 642
Db 1149 SVQITKFLIWIINDILVPAFQSENPQKQFVALYIOESLKCYGLSSSEMDNHKELYPN 1208
QY 643 DQATRAGLELKVLSFT--HPI--SFHSEFTESLACLKNDDEVAAE-----ALQ 690
Db 1209 EAKL--WEKNSVSKTITYPLTSSLYLAQSKREVVP--LKYPNSNFKGQYQIWKRRFTLD 1264
QY 691 IFKNTG-----SKIEEDPHIRSAALPVHL-----HKSCKGP----- 723
Db 1265 LTK-TGTENPHGVHSSLIREDGSLNPLPYISDIIIAKECTPYADLNGIIEF 1323
QY 724 -----ROAKYAIHCIAIFSSKE-----TOFAQIEFPLH---KSLDSPN 759
Db 1324 DSIFPCNLEGMNNLOVDSLRCQYESIFVFEYCKKMAEFKQNYSKLHGTFIIRKTKTN 1383
QY 760 LEHUTPLVTIGITALAPDOFAAPKSWAVATFYKDLMDRLP----- 804
Db 1384 M-----LRIIDEFLRTTPS-----DLAORSLTDSFERSALYLEOCY 1421
QY 805 -----GKTTKLWVPDEEV-----SPETWVIOAIKKMVRW----- 835
Db 1422 RQNPDRKQNGOLKLNQIITYEEIGDIDSLQVLTFTATGNLVSTEEIQLQSENKKLQD 1481
QY 836 ---LIGMKNHNSKSGTSLR-----LTTILHSDGDLTEQGIS--KPDMSRLAAGSA 885
Db 1482 CFNVLGKFSDDPKTTTMRMLKSMYDQLYQIISNSFSHSDQKISLSPDVK----- 1532
QY 886 IVKLAQEPCHIEITLBOYCALAIND--CYQVQVPAQKIH--KGLSRRLPLEY--- 939
Db 1533 -----EMYSIGLEANLEGNVQTLKKNVEQIESLRNIDREVLQYVIA 1576
QY 940 MAICALCAKDPVKERRAHAROCLVNINVRREYLKQHAASEKLSLLPEYVVPYTIHL 999
Db 1577 KALAIASMEDPLRQ-----KYINSRRLIGTNITTSKETTLLKQNLMLKMLSL 1627
QY 1000 AHPDPIV-KVODIEOLKDVKECLWFEILMAKNENNSHAF--IRRMVENIKQTKDAQSPD 1057
Db 1628 -YDLSFLSSAKDKEKSNNTIILDYRMERICADAEVPHNYILSMRKSFDQLKMNDEA--- 1682
QY 1058 DAKMNEKLYTCDAVMN-----IIMSKSTYSLE 1086
Db 1683 DADIGKTFITLAQIARRNARLDIASESLMHCL 1715

```

```

RESULT 10
US-08-870-693-15
; Sequence 15 Application US/08870693
; Patent No. 3866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.

```

```

; APPLICANT: Weinert, Ted A.
; APPLICANT: Pion, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCRI10798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2368 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: yeast MDC1 protein
;
; US-08-870-693-15

```

```

Query Match 2.5%; Score 178.5; DB 2; Length 2368;
Best Local Similarity 19.4%; Pred. No. 5.5e-05;
Matches 251; Conservative 188; Mismatches 477; Indels 377; Gaps 60;

QY 68 FLKHPGKD--VRLVACCLADIFRIYAEAPYTSFSDKLDIFMFTIRQLKGLDPTKSFQF 125
Db 526 FLSHNIDFSESLSGLIFSLHRFS---HFQPKLTDNGQINKSFK-LVQKCFMNS 580
QY 126 NRYEFLLENIMAVKSYNI--CFELEDSEIEFTOLYRTLFVJNNGHNQKVMHMYDLMS 183
Db 561 NRYELRL-----STRILPFIINSDSHNSEDEHTATILIKLQSKLPEVAKENVIAMTQ 633
QY 184 IICEDIVSOELDLTVLVNLPVPAHKN-----LNKQAYDLARALKRGAQIEPYITTT 235
Db 634 LTL--LTSNDVFDPLILKLDIFNSDDYSLRIMWTIQIKMMAKILKTPQQLSPILPV 690
QY 236 FFNOVLAMGKISISDSLSHVVDLILELYNIDSHLLS-----VLPOLEFKLSKNDNERL 290
Db 691 LLRQ---LGNLVE--RRVGFQNLIELLGYPSKTYLIDIFORYIIPYALIQKSD----- 739
QY 291 QVVKLAKMFGAKDELSAQNKP-----LMQCYLGR---F 322
Db 740 -VLSIAIKMGDQPSLNLQMKVNLKKNRQIFAVALYKGLSLDLETLFLNRPATF 798
QY 323 NDHVPRLKCVK-FASHCLMNHPLAKD-----LLEYLKVRSVSH--- 360
Db 799 DKGITAVLPDYKTLAEITKLYKNSVTKDASDENANMILCSLRFITNFEKDKRHSKY 858

```

```

QY 361 -----DPEAIR---HDVIVSYTAAKKDILLVNDHLLNFVREPTLK-----RWV 404
D 859 KNINMTDDDOQAQKQLQDNILGIFOVSSDIHDVGRITTYEKLKLVINGISLITAYP 918
QY 405 KREAMMGLAOYKRYALQSAAGDAKQAMIKDKLHIYONSIDRL--LVERIPA-- 460
D 919 KKSIIISALQI--SICLOTGIG--LKEVREAFRCNHLVRLHNLDELSTVIDSLAFI 973
QY 461 -----QYVPHNLETTERMKCLYLL-----YATLDLNAVKALNEMKCONLIR-- 503
D 974 LQKSEFNGKLRNIVYSIDTLTKKESDILLKLPYTTALVGPBELGILARDQAFARMV 1033
QY 504 HOVADLDLJKOPKTDASVAFSKVAVITRNLPD-----PGKAQ 543
D 1034 NKIRISTDIL--PIFANNLKS--SNKYVINQNDLIDIVYLRRKOTERSIDETPKKVGOTS 1089
QY 544 DFMKFTQVLEDEDEKIRKOLEVIVSPYSCQKQAEVREITTK-----LGNKOPTNP 596
D 1090 DITLVGLALLDTSKHFRLDKCEKCAKCSMIG--VLADYTKHEFKRTYSENEVYDLND 1148
QY 597 FLEKIRKLL--ERIPVHIDESIS--ALIKOVNKSIDSTADD-----EDEGVP 642
D 1149 SVGTIKELIWINIDILVPAFQWSENPSCOLEVALVIOESLKYCGLSESMDMNHKELYPN 1208
QY 643 DOAIRAGLELLKVLFT--HPI--SFHSAEFESLACLKMDDEKVAEA-----ALQ 690
D 1209 EAKL--WEKFNVSQKTIITPYLLSSLYLAOSWKEYVP--LKYPSNFEQGIQIWKRTTLD 1264
QY 691 IFKNTG-----SKIEEDPHIRSALEPYLH--HKSQKGP----- 723
D 1265 LK--TGTTENHPGHVFSLSLREDGSLNELLPYISLDITIKAKGPRYADILNGITIEF 1323
QY 724 -----ROAKYAIHCHIAITSSKE-----TOFAQIEPH--KSLDSPN 759
D 1324 DSIFTCNLEGMNNLQVDSLKCEYESIFRVEYCKWATERKQNSKLHGFTIIDTJTN 1383
QY 760 LEHLITPLVITIGITALAPQOFAAPMKSWATFIVKDLMDRLP----- 804
D 1384 M-----LIRIDEFLRTPS-----DLAKORSLETSPERSALYLEOCY 1421
QY 805 -----GKTTKMLVDEEV-----SPETVYKQAIKMWVW----- 835
D 1422 RQNHDKRNGOLKKNQIITYEEIIGDIDSLDGLVTRPATGNIVAKIELOQYSEWKLAD 1481
QY 836 ---LLGKNNHKSQSTLR-----LTTIHSDDGLTBOGKTS--KPDMSRLLAGSA 885
D 1482 CFNVLGFSDDPKTTTMLKSMYDHLQYLSQIISNFSHSSDGKISLSPDVK----- 1532
QY 886 IVKLAOPCYHEIITTEYOULCALAINDE--CYQVROVFAOKLH--KGLSRLRLPLEY-- 939
D 1533 -----EMYSIGLEAANLEGNVOYLKMWQVQIESLRNIDIREVLQYNIA 1576
QY 940 MAICALCAKDPYKERRAHOACLYKKNINVRREYIKQHAANSEKILSLPEYVVPYTHLL 999
D 1577 KALIASNEPDLFTQ-----KYIHNSEFRLGTGFTITSKRTTLLKQNLMLKHS 1627
QY 1000 AHPDQVY--KYODIEOULKQVKECLMEVLEIIMAKNENNSHAF--IRKMEVNIKQTFDAQSPD 1057
D 1628 -YDLSFSSAKDKFEYKSNMTTILDYRMRERIGADVPHVYILISMKSPDQKMKNEQA-- 1682
QY 1058 DAKMNEKLYTVCDVAMN-----IIMSKSTYSLE 1086
D 1683 DADIGKTFPTLAQIARANNARLDIASESLMHCLE 1715

```

RESULT 11
 US-08-290-731C-2
 : Sequence 2, Application US/08290731C
 : Patent No. 5843646
 : GENERAL INFORMATION:
 : APPLICANT: BOWTELL, David Douglas Lawrence
 : TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE

```

: TITLE OF INVENTION: SON OF SEVENTEEN (MSOS) GENE,  

: TITLE OF INVENTION: AND MSOS POLYPEPTIDES  

: NUMBER OF SEQUENCES: 15  

: CORRESPONDENCE ADDRESS:  

: ADDRESSEE: SUGHRUE, MION, ZINN, MACPHEAK & SEAS  

: STREET: 2100 PENNSYLVANIA AVENUE, N.W.  

: CITY: WASHINGTON  

: STATE: D.C.  

: COUNTRY: USA  

: ZIP: 20037  

: COMPUTER READABLE FORM:  

: MEDIUM TYPE: Floppy disk  

: COMPUTER: IBM PC compatible  

: OPERATING SYSTEM: PC-DOS/MS-DOS  

: SOFTWARE: PatentIn Release #1.0, Version #1.25  

: CURRENT APPLICATION DATA:  

: APPLICATION NUMBER: US/08/290,731C  

: FILING DATE: 17-OCT-1994  

: CLASSIFICATION: 435  

: PRIOR APPLICATION DATA:  

: APPLICATION NUMBER: PCT/AU93/00068  

: FILING DATE: 17-FEB-1993  

: PRIOR APPLICATION DATA:  

: APPLICATION NUMBER: PLO921/92  

: FILING DATE: 17-FEB-1992  

: ATTORNEY/AGENT INFORMATION:  

: NAME: KIT, Gordon  

: REGISTRATION NUMBER: 30,764  

: REFERENCE/DOCKET NUMBER: Q-36066  

: TELECOMMUNICATION INFORMATION:  

: TELEPHONE: (202) 293-7060  

: TELEFAX: (202) 293-7860  

: TELEX: 6491103  

: INFORMATION FOR SEQ ID NO: 2:  

: SEQUENCE CHARACTERISTICS:  

: LENGTH: 1319 amino acids  

: TYPE: amino acid  

: TOPOLOGY: linear  

: MOLECULE TYPE: protein  

: US-08-290-731C-2

```

```

Query Match 2.44; Score 176; DB 2; Length 1319;  

Best Local Similarity 19.24; Pred. No. 3.4e-05;  

Matches 252; Conservative 200; Mismatches 494; Indels 364; Gaps 68;  

QY 189 DTVSQELDVLVNVVPAHNNLKAQA--YDLAALLKRTQAI-----EPYIT 225  

D 174 DMFHQVEDINILSLDEPSTSGEOTYDLYKAFMAETROYIRELNLIKVREPVSN 233  

QY 236 FFNQVIMLGKTSISDSEHVFDDLILLYNIDSHLLSVLPOLFEKLSNDEBERLQVVKL 295  

D 234 -----SKLFSNDVENEFSTRIVDIHESVALLGHIEDTVEMTDEGSPH-----PL 278  

QY 296 LAKMFGAKDSELASQNKPLMQCYLGRPNDIHVRILECVKFSACL--MNHPLADLTE 353  

D 279 VGSCEFEDLABELA-----PDYSEYARDILRP-----GFHGHFLSQLSKPAAL-- 322  

QY 354 YLKVRSNDPEARLHDVIVSYTAAKKDILLVNDH--LNFVREPTLDRMR--VKREAM 409  

D 323 YLOSIGRGEFAOY--VLPRLILAPVYHCL--HYFELLIKOLEESSEDEDECKMQAI 377  

QY 410 MGLAQIYKRYALQSAAGDAKQAMIKDKLHIYONSIDRLVLE--RIFAQVAVPH 466  

D 378 TALLNV-----QSGMKTSKSLA-----KRLSSACRFYSQMKK 415  

QY 467 NLEETTERMKCLYLYATLDLNAVKALEMMKCONLRLHQYKDLLD--LIKOPTDAS 521  

D 416 OL-----AIKKNNEIOK--NIDQWEGKDIOGCCNEFTIMEGTLTVRG 454  

QY 522 VK-----AIRSKYAVITRN-----LPDPKQAQ-----DFMKFTQVLEDE--KIRKQ 562  

D 455 AKHERHIFPDGMICKSNHGQPRLPAGASSAEYRLKEKFPMK--VOINDKDTSEKHA 513

```



```

Db 577 OEEKEEQMRLEPSAEVYFFAEEDSEENILFEENVOPKAGIPILIKAGVILKILERLTMYMA 636
Qy 667 -----AEFF-----ESILACL-----KMDKAVADALQIFKNTGSKIEEPPIHRS 708
Db 637 DPNFVFRTLTYSFCRPOELLSTLIERFEIPEPEPEARIALENDQOISALKFRK 696
Qy 709 ALLPVLHHSKKSGPPROAKYAIIHCIAIFSSKETOFQOIFEPRLHKSIDPSNLHLITPLV 768
Db 697 EYI-----QPVOLRVLANCRHW-----EHHFVD-PE-----RDADLLQMEEFIG 736
Qy 769 TIGHIALLAPDQEAAPKMSVAT--FIVKOLMNDRLPGKKTTLWNPDEEVSPEPMVK 825
Db 737 TVRKGAM-----KKWVESITKIIORKKIARDNGPGHNIT-----FQSSPP-- 777
Qy 826 IQAIKMMVRLGLGKNNHSG--TSLRLLT-----TILHSD-----GD 863
Db 778 -----VEMHI-----SRGHITETDILLHPLEIAROTLLESIDLRAVOPSELVGS 824
Qy 864 L--TEQKISKPDMSRLRLAAGSAIVKLAOEPCEYHEITLLEQYOLCALAINDECYQYRV 921
Db 825 VMTREDEKEINSPN--LKMIHRTNLTLMFEKCIVENLENLE--RVAVSRITIELQY 878
Qy 922 PAQKLHGLSLRLPLEYMAICALCAKDPV-----KERRAHROCLVKNINVRREYLK 974
Db 879 FOE-----LNNFNGVLE--VSAMNSSPYRLDHTFEQILPSROKKILIEAHLESEDHYK 930
Qy 975 OHAAVSEKLSLLPEYVVPYTIHL-----AHDPDVYKVDIQLDVKECIMF----- 1023
Db 931 KYLA--KLNSINPCVPFGIYVLTNLKTEEGNPEVL-----RHGKELINFSRRR 980
Qy 1024 VLEILMAKNENNSHAFIRKMWENIKOTKDAQPPDAMNEKLYTCDVAMNIIIMSKSTTY 1083
Db 981 VAEITGLIQYQNGPYCLARVBPDIKRFENLNPNGNM--EKELF-----DYLENNGLEI 1033
Qy 1084 SLESPKQVLPARFFTOPDKNFSTKNYLLPREMKSFTTPGKPKTTNVLGAVNKPL--SS 1140
Db 1034 EPRAPKP--LP-RP-----PKKYSYPLKSPGVAPSPNRP-----GTMRHPTPLQOE 1076
Qy 1141 AGKOSOKSSMETVSNASSSNSP-----SPGRIGRLDSSSEM-----DHSENDYDMS 1191
Db 1077 PKRISYRIPESEIESTASAPNSPRTPLTPPASGTSNTDYCVFSDSHASFPFHSA 1136
Qy 1192 PLPGKSKDKRDDSDLVSELEKPPRGKRTPYTQOEKL--GMDLTKLVOEQ-----K 1242
Db 1137 SVSISLSKGTDEVPPVPVPPRRRPSAPAESPSKIMSKHLSPALIPROQTSKAYS 1196
Qy 1243 PKGQBRKRKCHTASDEDOQWPEEKRLKEDILENEBONSPPKKKRG----- 1291
Db 1197 PRYSISDRTSISDPRESPLLPREPVRTPDVFSSPLHLQRPPLGKSKSDHGNAFPNNSP 1256
Qy 1292 ---RPPKPLGGGTPEKEPTMK-----TSKKGSKKSKSPAPAREEHEE 1331
Db 1257 SPFFPPPP---QTPSPHGTRRHLPSPLTQMDLHSLAGPVPVPROSTSQ 1303

```

RESULT 13
US-08-973-462-8

Sequence 8, Application US/089734628

Patent No. 6191270

GENERAL INFORMATION:

APPLICANT: DRUTHE, PIERRE

APPLICANT: DAUBESIES, PIERRE

TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

FILE REFERENCE: 0660-0125-0 PCT

CURRENT APPLICATION NUMBER: US/08/973, 462B

CURRENT FILING DATE: 1998-02-06

EARLIER APPLICATION NUMBER: PCT/FR96/00894

EARLIER FILING DATE: 1996-06-12

EARLIER APPLICATION NUMBER: FR 95/07007

EARLIER FILING DATE: 1995-06-13

NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

Query Match 2.4%: Score 174; DB 3; Length 1786;
Best Local Similarity 17.0%; Pred. No. 8, 1e-05;
Matches 203; Conservative 192; Mismatches 398; Indels 402; Gaps 48;

```

Qy 101 DKLDIPM-----FITROLGLEDTSQPQFNRYFYLENTA--WKSYNICELED 149
Db 838 EEIKDSILNEIEYKENVVTTILENVBETTAESVTFNSILIEIENFTINDTIEERLEE 897
Qy 150 SNEFTQLYTIPSVINNGHNOKVHIMVNDLMSIICEGTYQOELLDY-----LVN 202
Db 898 LHE-----NVLSALENTQSEEEKKEVIDIEV--KEEVATTLIETVEQAEEKSANT 948
Qy 203 LVPAHKMLNKQAVDLAKALKRTQOAIPEYITTFEFOVLMIGKTSISDLSEHFDLITEL 262
Db 949 ITEIFENLEENAVESN-----ENVAENLEKINETYENTVL-----DKVEET 989
Qy 263 YNIDSHLLSVLPOLFEKLSNDNEERLQYVYLLAKKAGKADSELASQNKPLQCYLGRF 322
Db 990 VEISGESLENNEKOKAFSEIFDVKIQ--ENLLTGFRSIEISIVYIQSEKVD-----L 1043
Qy 323 NDHIVPIRECVAFASCLMNHDP-----LAKDLEYLKVSHDPEEAIIRDVIYSIV 375
Db 1044 NENVVSSILDNIEEMKGLNKLLENISSTEGVQETVEH-----VEQNVYVDV- 1091
Qy 376 TAAKDLILVNDHLNFRERTLDKRWVRKEAMMAGIAQYKKYALQSAAGKDAKQIAW 435
Db 1092 -----DVPAAKQDQGLGLNZA-----GGLKEPFNLEDYFK----- 1122
Qy 436 IKDKLHIYYQNSIDRLVERLIFAQYVPHNETTERMKCLLYATLQNALNVALNEM 495
Db 1123 -----SESDVITVEIKDE--PVQKE-----VEKEVYSIIEEM 1153
Qy 496 WKQCNLRHVOYKDLIDLKQPKTDASVKALFSKMYVITRNLPRGKADQPMKFTOYLED 555
Db 1154 -----ENIVDVLEEEKEDLDTKMDAABESIEISSDKKEEESIKDKEDVSLVEE 1206
Qy 556 -----DEKIRKQLEVLVSPTCSCQOAGCVRETKKLGKPKQPTNPF--LEMIKFLERI 608
Db 1207 VQNDMDSEYKVELKEMMEELKQAVE--INDTTSILIBETQELNEVEADLK----- 1259
Qy 609 APVHIIDYESIALIKOVNKSIDGTADDECEGVPDQAIRAGLELLKVLSTFHPISFSAE 668
Db 1260 -----DMEKLEKELEKALSEDSKEIIDAQD-----TLEKVIIEEHDT-----T 1298
Qy 669 TFEISLACIMKMDKVAEAA-----IQIFKNTGSKIEEDPPIHRSALLPVLNH 716
Db 1299 TLDEVVELKQVEEKIEKVSDLKDLIEDLIKVEKIELESELEDYKELKTJETDILIEE 1358
Qy 717 KSKKGPPOAKYAIHC-----IHAIFSSKETQFAQIPEPLHKSIDSNLEHLITPL 767
Db 1359 KKEIEKHFEKEFEAEAEIENDLEADLIKVESLSLEVEEKRLIEVHEIKE--EVEHITS-- 1414
Qy 768 VTIGHIALLAPDQ--AAPKMSVATFIVKOLMNDRLPGKKTTLWNPDEEVSPEPMVK 825
Db 1415 -GDANHIGLEEDDLEEVDDLGSLIDMLKGMOMELGDW--DKESL-----EDVYTKIGER 1465
Qy 826 IQAIKMMVRLGLGKNNHSGSTSLRLTLTILHSDDLVBQGISKPDMSRLRLAAGSA 885
Db 1466 VESLKDVLSSALGDEBQOMKTR----- 1487
Qy 886 IVKLAOECYHEITLLEQYOLCALAINDECYQVROFAOKLHKSRLRLPLEYMAICAL 945
Db 1488 --KKAQRPKLEEVLLKE-----EVAEEKKKITK--KKVRDIK----- 1522
Qy 946 CAKDPVERRAHAROCVKNINVRREYLKOHAAVSEKLSLLPEYVVPYTIHLAHDPDY 1005

```

Db 1523 -DKEKDE-----IYEVEMKDEDEED--VEED-----IEEDIEE 1554
 Oy 1006 VKVDIOEQ--LKDYECIMFVLEILMAKNENNSHAFIRKMNENIKOTDAGCPDDAKANE 1063
 Db 1555 DKVEDIDEDIDEDIGEDKDEVIDLTVOK-----EKRIEKVAKK-----KLEK 1598
 Oy 1064 K-----LYTVCDVAMNII-----MSKSTYSLESPKD----- 1090
 Db 1599 KVEBVSGLKKNHVDVEMKYVOKIDKEVDKEVSKALESKNVDYTNVLKONODEFSKVNFK 1658
 Oy 1091 -----PVLPA-----RFFP-----OPDKN----- 1104
 Db 1659 KYKVAAPFISAVNAFAFAYVGFPPFLSSCVTIASSTYLLSKYDKIINKKERPFYSF 1718
 Oy 1105 -----FSNTKNTLPPEMKSFPTPGKPTTNVLGAVNKPILSSAGKOSOTSSRMETV 1155
 Db 1719 VEDIFKNLHYL--QOMKEKFS--KEKNNNVIEVTNKAERKGNVQVTNKTETTKV 1770

RESULT 14
 US-08-588-985-2
 ; Sequence 2, Application US/08588985
 ; Patent No. 5777094
 ; GENERAL INFORMATION:
 ; APPLICANT: MICHIOYUKI MATSUDA et al.
 ; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Menderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/588, 985
 ; FILING DATE: January 19, 1996
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1865 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL:
 ; ANTI-SENSE:
 ; FRAGMENT TYPE:
 ; ORIGINAL SOURCE: spleen cell of homo sapiens
 ; US-08-588-985-2

Query Match 2.4%; Score 172.5; DB 1; Length 1865;
 Best Local Similarity 17.6%; Pred. No. 0.00012;
 Matches 258; Conservative 212; Mismatches 441; Indels 553; Gaps 72;

Oy 6 TRNDGKITPPGVKEISDKISKEEMVRLKMNVTFTMDMDQDSPEEKELYNLNLHLAS 65

Db 803 TIVNDVKLVPPD--KELS-----KMFTEFINLVPMGLTICKLCLIEI--VHS 847
 Oy 66 DPEFLKPECKDVRL-----LVACC--LADIFRI-YAPAPPTSPDKLDI 106
 Db 848 DLFQHDCEREILLPMPTDQLKYLHERQDLEACCOLLSHILEVLYRKDVGPQO---RHV 903
 Oy 107 FMFITROLKGLDEFPSPFNRYFYLEN-IAMVSYNCFELEDSENEIFQIYRTLFSEVI 165
 Db 904 QIMEKRLRYVNRVYI-SMGDSSELIGNVACMTA--ILRQMEDYH--YAHLIKTE----- 954
 Oy 166 NNGHNQVYMHMVD-IMSSIIICEGDTVQOEI--LDTVLVNLVPAKKNLNKOAYDAAKALL 222
 Db 955 -----GKMRIVDVPELMETFLFMFKNLICKNNYPPEGMVIMNVQ-----NKFL 997
 Oy 223 KRTAQAIEPYTTTFEFOVLMIGKTSISDLSEHVEDLILELYNIDSHLLSVLPOLFEKIK 282
 Db 998 R-----AINQYAD-----MLNKFLDQAN--FE--LQIWNVYFHLAVAFLETSIQLE 1041
 Oy 283 SINDNERLOVYKLLAKMGAKDSELASONKPLMOCYLGFRPNDIHYPIRLCEGVKASHCLM 342
 Db 1042 NFSSAKR--AKILNK-YGDMKRQIGFEIRDMVY-NLGQHK-----IKFL----- 1081
 Oy 343 NHPDLAKDLYELKVRSHDPEAIRHDI-----VSIVTAARKDILLVNDHL 389
 Db 1082 --PENVGPILEMTLI-----PETELRKATIPPIFFDMQCEPSTSFQFEMEITKDHE 1135
 Oy 390 LNFVREERTLDKRWVRKRAMMGIAQYKVALQSAAGDAKQAIAMIKDKLLHYONSI 449
 Db 1136 VEGGRG--DEQYVLFDKI--LEHCRRHKHYLAKTGTFVFLVY----- 1175
 Oy 450 DDRLIVERIFPAQVYVPHLETTERKCLYLYATLIDLANVALNEMKCOYLHROYKDL 509
 Db 1176 --RLMERLDDYRTIMDENKENRMS--TAVNVLNFKYEIEREEMVIRLYK-L 1223
 Oy 510 LDLIKOPRTDASVKAIFSKVWVITRNLPDPKADQFMKFTVOLEDDIKIKOLEVLVSP 569
 Db 1224 CDLHKE--CDNYTEAATYLL-----NAKLKMSDY----- 1253
 Oy 570 TCSCQKQAGCVREITKKLGNPKOPTN-----FLEMIKFLERIAVPHIDTESISALI 622
 Db 1254 -----CVANLQBRG--YQATQOQLKEQLYQELIHF----- 1284
 Oy 623 KOVNSIDGTADDEDECVPTQOATRAGIELLKVLSTFPIPSHAEFPESLACIKMDE 682
 Db 1285 -----DKGKMEEAIALKEL-----AEQYENM--PDYE 1312
 Oy 683 KVAE--AAOIIFKNTGSKIE-----EDPHIRGALLPVLLHKKSKGPPROAK 727
 Db 1313 QISELKKQAOFYENIYVIRPKRDYRANGYGGGFPFLRG--KVFIYRKEVEYPRPD- 1369
 Oy 728 YAIHCINAIFFSK-ETQFAQIFERLYHKSLSDSNEN-----LITPLVTIGHIALLA 777
 Db 1370 -----FEARILITOPRPAEKKMTTSPPDGDIKNSPGOYIQCFYVAPKIDL----- 1413
 Oy 778 PDQFAAPKSNVAIF-----YKDLMLNDRLEKKTTKMLVPRDEEVSPEYVAKIQ 827
 Db 1414 PPKFHRPVSQIIVSFYRVEVQREYSPRIKRGKNDPNEPANNMI--ERTIYTTAVKLP 1471
 Oy 828 AIKMMVRYLGMKNHNSKSGTSLRLTLTIHSQDLTEQGIKSKIPDMSRLAAGSAIV 887
 Db 1472 GI---LRF-----EVSVMVEISPLENA----- 1493
 Oy 888 KLAOEPCHETITLEOYOLALAINDECYOVROVFAOKLHGLSRLRLPEYMAICALCA 947
 Db 1494 -----IETQDLTNDKIN-----SWVOQHDDPSLPINPLMLNGI 1529
 Oy 948 KDP--VKERRAHAROCLYKKNINVRREYLYKONAAVSEKLLSLPEYVVPYTTNLLAHDPYV 1006
 Db 1530 VDPAVVMGFAFYEKAFFTD-----RYLOEHPHAHEK----- 1560
 Oy 1007 KVODIEOLKDVKECLM---FVLEIIMAKNENNSHA---FIKRVENIKOTDAGCPDDAK 1060

```

Db      1561 ----IEKLIKDL--IAWQIFLAEGIRIHODKYTEALRPHERMEACFKOLKEKV----- 1608
QY      1061 MNEKIYTVCDVAMNIIMSKSTYTSLESPPKDPVLPAPFFTOPKNSNTNLYLPPEKKSFF 1120
Db      1609 ---EKEYGV---RTIPSSLDROGRSPRSNV---RSFTMPSS----- 1641
QY      1121 TPGRKPTNVLGAVNKPLSSAGKQSQTKSRMETVSNASSSNPSSPGRIKGRLDSSEND 1180
Db      1642 -----SRPLSLVA-----SVSSLSDSPTSPSRG----- 1663
QY      1181 HSENDYTWSSPLPGK---KSDKRDDSLVNSELEKPPGR-----KTPVT 1223
Db      1664 ---SGFALEPLPLPKMHSRQDKLDKDLKEKPKKKEKRSKHQEIFEKEFKPTDIS 1720
QY      1224 -BOEELGMDLTKIVQEQPK-----GSQSRKRGHTASRDEQWPEEKRLK----- 1271
Db      1721 LQOSEAVILSETISPLRPQPSQVNMVNTGSEKRSVSPSSSQQTPTPTPRALSTS 1780
QY      1272 -----EDILENEDEONS-----PRKGRGRPPRP 1296
Db      1781 MGSLELNGMTGADVADVPPLPLKGSVADYGNLMENODLSPPTPPPPHQRHLPP-P 1839
QY      1297 LGGCTPKPEPTMTSKSGSKSKSG 1320
Db      1840 LPSKTPPPPP-KTYRKQTSVDSG 1862

```

```

RESULT 15
US-08-971-988-2
; Sequence 2, Application US/08971988
; Patent No. 5786461
; GENERAL INFORMATION:
; APPLICANT: Michiyuki MATSUDA et al.
; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,988
; FILING DATE: 17-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/588,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1865 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: spleen cell of homo sapiens

```

```

US-08-971-988-2
Query Match      2.4% Score 172.5; DB 1; Length 1865;
Best Local Similarity 17.6% Pred. No. 0.00012;
Matches 258; Conservative 212; Mismatches 441; Indels 553; Gaps 72;

QY      6 TPTNDGKITYPGVEKISDKISEMVRRLKAVNTFMDMDQSEEEKELYNLALHLAS 65
Db      803 TIVNDKYLVEDP--KELS-----KMEFERILVPMGLTIQKLYCLIEL-VHS 847
QY      66 DEFLKHPGKDVNL-----LVACC--LADIFRI-YAREPRTSPDKADI 106
Db      848 DLFTQDCEKREILPMTDQKLYHLERQEDLEACCOLSHILEVLVRKDDGPTQ---RHV 903
QY      107 FMEITROLKGLDTSQPFNRYEVLLEN-IAWVSNICEFELEDSNEIPTOLRYTLFSVI 165
Db      904 QIIMEKLLTVNRITYI-SMGRSEILGNVACMTA--ILROMEDYH-YAHLIKTF----- 954
QY      166 NNGHNOKVIMHVD-LMSSIIIGDVTSGQL--LDFTVLNVLPAHKNLKOAYDLAKALI 222
Db      955 -----GKMRDGVVDLFMEFTIMEKNLIGKNVYRFDVIMNVQ-----NKVEL 997
QY      223 KRTQAQIEPYITTFEVOVLMLGKTSISDSIENVPDLILELYNIDSHLLSVLPQLEFKL 282
Db      998 R---AINQYAD-----MLKKKFLDQAN---FE--LDLMNNYFLAVAFLOESLOLE 1041
QY      283 SNDNEERLOVAVKLAKMFGAKDSELASQKPLMOCYLGFRNDIHVIRLECYKFAHSCLM 342
Db      1042 NFSSAKR---AKILNK-YQDMRQIQEFLRDMYV-NLGQHK-----IKFI----- 1081
QY      343 NHPDLAKDLEYLKVRSHDPEEARHDI-----VSIVTAARKDILLVNDHL 389
Db      1082 --PEWVGPILEMTLI---PETELRKATIPRIFDMQCSFHNSTRSQFMENEITIKLDEH 1135
QY      390 LNFVREERTIDKRRVAKEMAGLOAYKKYVALQASAGKRAAQIAIKIKLHIYYQNSI 449
Db      1136 VEEGRG--DEQKVLVDKI--LLECRNHYLAKGFEVFLV----- 1175
QY      450 DORLVERIFAQVMVHNLEETTERMKCLYLYATLDLNAVKALENMKCONLIRHQVKDL 509
Db      1176 ---RLMERLLDYRTIMHDEKNRMSK-----TVNVLNYYKEIEREMVIRIYK-L 1223
QY      510 LDLIKQKTDASVKAIFSKVMYITRNLPDPAQODEFMKFTVOYLEDEKIRQOLEVLVSP 569
Db      1224 CDLHKE--CDNTEAAYTLIL-----HAKLLKSEDEV----- 1253
QY      570 TCSCKQAEQCVREITKILGNKQPTNP-----FLEMIKFLLEIARPHIHIDESTALI 622
Db      1254 -----CVAHLTQRDG--YQATTQOGLKEQLYQELIHYF----- 1284
QY      623 KOVNSIDGTADDEGCVPTDOAIRAGLELLKVLSTNPLSPHSAETFESSLACLKMDDE 682
Db      1285 -----DKGMMEEBALGKEL-----AQYENEM-----FDVE 1312
QY      683 KVAE---AALQIFKNTGSKIE-----EDFPHIRSAALLPVLHNSKKGPPRQAK 727
Db      1313 QISELLKKAQAFENIVKVIYRPKDYEAUYGQGPFTLRG--KVFIYRKGXEYBRPD- 1369
QY      728 YAIHCHIAIFSSK-ETQPAQIIEPRLKSLDPSLNEH-----LITPLVYTIQHALLA 777
Db      1370 -----FEARLLTFPNAEKMKYTPSPGDGDKNSPOQYIQCFTVKRKLDE----- 1413
QY      778 PDGFAAPKSWATF-----IVKDLMDRLPGKKTRLAVPDEVSPEYWKIO 827
Db      1414 PKFHRPVSQIVSFVRVNEVQFEYSRPIRKEKRPDNEFANMVI--ERTIYTTAVKLP 1471
QY      828 AIRMAVRMLGKNNNSKSGSTYLRLLTTLILHSDGLTDOGKISKDMSRLRLAASAIY 887
Db      1472 GI---LRWF-----EVKSVFVEISPLENA----- 1493
QY      888 KLAQEPCEYHEIITLEQYQALAIINDECYQAVRQAKLKGISRLRPLEYVATGALCA 947
Db      1494 -----IETMOLTNDKIN-----SMVOQHLDPSLPINPLSMULNGI 1529

```

```
QY 948 KDP-VKERRAHAROCVLVKNINVRREYLKQHAAVSEKLSLLEPYVVPYTIHLHLPDXY 1006
Db 1530 VDPVAMGSPANERKEFTD-----RYLOEHPAHEK----- 1560
QY 1007 KVODIEQLKDYKECIAW---FVLEILMAKNENNSHA---FIRKVENIKOTKDAQCPDDAK 1060
Db 1561 ----TEKIKDL-IAMQIPFLAEGIRINGDKYTEALRPFHERMEACFKOLEKV----- 1608
QY 1061 MNEKLYVCDVAMNITMSKSTYTSLESPPKDPVLPARFETQDPDKNFNTKNYLPPEMKSEF 1120
Db 1609 --EKEYGV-----RIMPSSLDRRGSRPRSMV---RSFTMPSS----- 1641
QY 1121 TPGKPKTINVLGAVNKKPLSSAGKOSQTKSSRMEYTSNASSSSNPSSPGRIKGRLDSEMD 1180
Db 1642 -----SRPLSVA-----SVSSLSDSTPSRPG----- 1663
QY 1181 HSENEYTMSSPLPGK---KSDKRDSDLVRSLEKPRGR-----KTPVT 1223
Db 1664 ---SDGFALPEPLLPKMHRSQODKLDKDDLEKEKKDKKKKRNKSKHQEIFEKEFKPTDIS 1720
QY 1224 -EQEKKLGMDLTKLVQEQPK-----GSQSRKRGHTASESDQOWPEEKRLK---- 1271
Db 1721 LQOSEAVILSETISPLRPQRPKSQVMNVIGSERRFSVSPSSQOTPPPVTPRAKLSFS 1780
QY 1272 -----EDILENDEONS-----PPKKGKRGRRPKP 1296
Db 1781 MOSSLLENGMTGADVADYPPPLPLKGSVADYGNLMEODLGLSPTPPPPPHQRHLP-P 1839
QY 1297 LGGTPPKDEPTMTKSKGSKKSG 1320
Db 1840 LPSKTPPPPP-KTTRKOTSYVDSG 1862
```

Search completed: September 24, 2003, 20:21:59
Job time : 35 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 16:28:13 ; Search time 59 Seconds
(without alignments)
3742.181 Million cell updates/sec

Title: US-09-512-581B-2

Sequence: 1 MAHSKTRTNDGKITYPGVK.....QKGRGRPSKTPSPSQPKNV 1391

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1980.DAT *
2	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1981.DAT *
3	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1982.DAT *
4	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1983.DAT *
5	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1984.DAT *
6	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1985.DAT *
7	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1986.DAT *
8	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1987.DAT *
9	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1988.DAT *
10	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1989.DAT *
11	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1990.DAT *
12	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1991.DAT *
13	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1992.DAT *
14	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1993.DAT *
15	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1994.DAT *
16	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1995.DAT *
17	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1996.DAT *
18	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1997.DAT *
19	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1998.DAT *
20	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA1999.DAT *
21	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA2000.DAT *
22	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA2001.DAT *
23	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA2002.DAT *
24	/S1D1/gcgcdat4/geneseq/geneseqp -emb1/AA2003.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	719.3	100.0	1391	21	AAI94702		Human androgen shu
2	4642.5	64.5	1297	24	ABG72803		Human tumour supre
3	2999.5	41.7	919	23	ABP64685		Human protein SEQ
4	2715.5	37.8	854	21	AAAB01382		Neuron-associated
5	1982.5	27.6	1218	22	ABMB7869		Drosophila melanog
6	839	11.7	217	22	AAU278827		Human full-length
7	758.5	10.5	363	21	AABS5462		Lung cancer associ
8	709	9.9	182	23	ABBS4560		Human euchromosom
9	640.5	8.9	333	22	AAAB9496		Human protein sequ

10	599	8.3	165	22	AAU01759	Human secreted protein
11	591	8.2	147	22	ABH89260	Human polypeptide
12	513	7.1	101	21	AAV55247	Human 5' ESR related
13	467	6.5	92	21	AAU02811	Human
14	276	3.8	91	22	AAU27999	Human secreted protein
15	247.5	3.4	233	22	AAH95449	Human contig polyp
16	232	3.2	268	22	AAH40883	Human protein sequ
17	231.5	3.2	2835	23	AAU75100	Human polypeptide
18	231	3.2	2663	22	AAH39097	Novel signal pathw
19	229	3.2	2633	22	ABG06505	Human polypeptide
20	228.5	3.2	1643	22	ABH66811	Novel human diagn
21	225	3.1	1743	19	AAH98879	Drosophila melanog
22	218.5	3.0	1026	22	AAH78825	H. pylori GHPO 175
23	218	3.0	1033	22	AAH79809	Human protein SEQ
24	210.5	2.9	2047	23	AAU75883	Human protein SEQ
25	208.5	2.9	5303	22	ABH67866	Human adhesion mol
26	208	2.9	364	22	ABU52998	Drosophila melanog
27	208	2.9	628	22	ABU52994	Human testes-deriv
28	208	2.9	1183	22	ABH58769	Human testes-deriv
29	207	2.9	1428	22	ABH68557	Drosophila melanog
30	206	2.9	2243	22	ABH84884	Drosophila melanog
31	206	2.9	2440	22	ABG20278	Murine protein, SEQ
32	205.5	2.9	2230	22	ABU07445	Novel human diagn
33	205	2.8	2349	24	AAO16359	Human translocated
34	205	2.8	2415	22	ABG20279	Human protein
35	203.5	2.8	1881	23	ABP73809	Novel human diagn
36	203	2.8	622	22	ABU52993	Candida albicans e
37	201.5	2.8	3248	22	ABH93975	Human testes-deriv
38	200	2.8	1278	22	ABH71882	Kinetochore protei
39	199.5	2.8	699	22	AAH79156	Drosophila melanog
40	199.5	2.8	718	22	AAH60140	Human protein SEQ
41	197.5	2.7	1530	22	AAH84882	Human protein SEQ
42	197.5	2.7	2266	22	AAH84885	Human protein, SEQ
43	195.5	2.7	1150	22	ABH59129	Human protein, SEQ
44	195	2.7	2519	22	ABG16636	Drosophila melanog
45	194.5	2.7	2482	16	AAH28826	Novel human diagn
						Human mltosin. Ho

ALIGNMENTS

RESULT	1
AA94702	
ID	AA94702 standard; Protein; 1391 AA.
XX	AA94702;
AC	
DT	01-DEC-2000 (first entry)
XX	
DE	Human androgen shutoff gene 3 (AS3) protein sequence.
XX	
KM	Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KM	chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer
KW	diagnosis; treatment; cytostatic; human; ss.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	419..459
Region	/label= "Subdomain I"
FT	/note= "Hank's conserved region"
FT	419..425
Region	/label= "Beta strand 1"
FT	426..433
Region	/label= "Mg-ATP binding loop"
FT	453..459
Region	/label= "Beta strand 2"
FT	472..482
Region	/label= "Subdomain II"
FT	/note= "Hank's conserved domain - beta strand 3"
FT	489..496
Region	/label= "Subdomain III"
FT	/note= "Hank's conserved region - alpha helix C"
FT	

FT Region 509..516
/label= "Subdomain IV"
/note= "Hank's conserved region - beta strand 4"
FT 525..548
/label= "Subdomain V"
/note= "Hank's conserved region"
FT 525..532
/label= "Beta strand 5"
FT 540..548
/label= "AlIha helix D"
FT 554..567
/label= "Subdomain Via"
/note= "Hank's conserved region - alpha helix E"
FT
FT
PN WO200050454-A1.
XX
XX 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-US04732.
XX
XX 24-FEB-1999; 99US-0121461.
XX
XX (TUFT) TUFTS COLLEGE.
XX
XX Soto AM, Sonnenschein C, Geck P, Szelei J;
XX
XX WPI; 2000-565451/52.
XX
XX N-PsDB; AAA28051, AAA28052.
XX
XX
PT New human androgen-induced tumor suppressor cDNA sequence termed
PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT diagnosis and treatment of prostate cancer -
XX
XX
PS Claim 2: Fig 1: 152pp; English.
XX
XX This invention relates to a human androgen-induced tumour suppressor cDNA
CC sequence termed "androgen Shutoff Gene 3" (AS3). The AS3 gene is located
CC on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
CC proliferation and use as a marker for the efficient diagnosis and
CC treatment of prostate cancer. The invention includes AS3 cDNA and protein
CC sequences, a vector comprising the cDNA sequence, a host cell transfected
CC with the expression vector, and a method for producing an AS3 polypeptide
CC comprising culturing the transfected cells. AS3 has cytosstatic activity,
CC and acts to suppress cell proliferation. The AS3 gene is useful as a
CC marker for the efficient diagnosis and treatment of prostate cancer. The
CC AS3 nucleic acid molecule can be used as a source of antisense agents for
CC sequence specific modulation of gene expression. The AS3 protein may be
CC used in the treatment of disorders caused by aberrant modification or
CC mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
CC or aberrant post-translational modification of the AS3 protein. This
CC sequence represents the human AS3 protein sequence.
CC
XX
SQ Sequence 1391 AA:

Query Match 100.0%; Score 7193; DB 21; Length 1391;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAHSKRRTNDGKITYPPGYKEISDKISKEPMVRRLKMYVTFTMDMODDSEEEKVLTNLA 60
DB 1 MAHSKRRTNDGKITYPPGYKEISDKISKEPMVRRLKMYVTFTMDMODDSEEEKVLTNLA 60
OY 61 LHLASDFLKHPEKDVRLVAVACCLADIFRIYAPAPAYTSPDKLKDIFMFTTRDLKGIEDT 120
DB 61 LHLASDFLKHPEKDVRLVAVACCLADIFRIYAPAPAYTSPDKLKDIFMFTTRDLKGIEDT 120
OY 121 KSPQFNRYFYLLFNIMAVSYNTCFLEBDSNEIFTOLYRTLSVINNGHNOQKYHMAVDL 180
DB 121 KSPQFNRYFYLLFNIMAVSYNTCFLEBDSNEIFTOLYRTLSVINNGHNOQKYHMAVDL 180
OY 181 MSIIIEGDTVSQELLDIVLVNLPVPAHKNLNKOAYLAKALLKRTQOATIEPYTTTFENOV 240
DB 181 MSIIIEGDTVSQELLDIVLVNLPVPAHKNLNKOAYLAKALLKRTQOATIEPYTTTFENOV 240

OY 241 LMLGKTSISDLSEHVEDLILELYNIDSULLSVLPQLEEKLSNDNEERLQVYKLLAKMF 300
DB 241 LMLGKTSISDLSEHVEDLILELYNIDSULLSVLPQLEEKLSNDNEERLQVYKLLAKMF 300
OY 301 GAKDSLASQNKPLMOCYIGRPNIDHVPTRLECYKASGLNMHPDLADLEYLKVRSR 360
DB 301 GAKDSLASQNKPLMOCYIGRPNIDHVPTRLECYKASGLNMHPDLADLEYLKVRSR 360
OY 361 DPEEARHVDIVSYVAAKKDILLVNDHLINPREETLKRMRVRKEAMMGLAQYKKTA 420
DB 361 DPEEARHVDIVSYVAAKKDILLVNDHLINPREETLKRMRVRKEAMMGLAQYKKTA 420
OY 421 LOSAAGKDAKQIAWIKDKLILHIYQNSIDDRILVRIEFAQYVWPNHLEETTERMKCLYTL 480
DB 421 LOSAAGKDAKQIAWIKDKLILHIYQNSIDDRILVRIEFAQYVWPNHLEETTERMKCLYTL 480
OY 481 YATLDINAVKALNEMKCONLRHOKYKDLIDLKOPTASVKAIFSKYVWITRNLPDYG 540
DB 481 YATLDINAVKALNEMKCONLRHOKYKDLIDLKOPTASVKAIFSKYVWITRNLPDYG 540
OY 541 KAADFMKKFTQVLEDEDEKIRKOLEVYSPSCQKQEGCVREITTKLGNPKOPTNPELM 600
DB 541 KAADFMKKFTQVLEDEDEKIRKOLEVYSPSCQKQEGCVREITTKLGNPKOPTNPELM 600
OY 601 IKFLERIAPVHIDTESISALIKQVNSIDGTADDEGVPTDQAIRAGIELLKYLSFTN 660
DB 601 IKFLERIAPVHIDTESISALIKQVNSIDGTADDEGVPTDQAIRAGIELLKYLSFTN 660
OY 661 PISFHAETFPESILACLKMDKEVAAALQIFKNTSTKIEEDPHIRSAALLPVLHKSKK 720
DB 661 PISFHAETFPESILACLKMDKEVAAALQIFKNTSTKIEEDPHIRSAALLPVLHKSKK 720
OY 721 GPPROAKVYAIHCHAFSSKETQFAQIFPIKHSIPSNLEHITPLVMTGHTALAPPO 780
DB 721 GPPROAKVYAIHCHAFSSKETQFAQIFPIKHSIPSNLEHITPLVMTGHTALAPPO 780
OY 781 FAAPKSWVATFIVKDLNMDRLPGKRTKLWVPDEVSPEYVWKIQAIKMYRWLLGKM 840
DB 781 FAAPKSWVATFIVKDLNMDRLPGKRTKLWVPDEVSPEYVWKIQAIKMYRWLLGKM 840
OY 841 NNSKSGTSLRLTLTILHSDGDLTEOGKISKDMRRLAASALVTKLAOECEYHEIT 900
DB 841 NNSKSGTSLRLTLTILHSDGDLTEOGKISKDMRRLAASALVTKLAOECEYHEIT 900
OY 901 LEOYQCALAINDECTQVROVFAQKHLKLSRLPLEYWAICALCAKDPVKKRRHARQ 960
DB 901 LEOYQCALAINDECTQVROVFAQKHLKLSRLPLEYWAICALCAKDPVKKRRHARQ 960
OY 961 CLVKNINVRREYLKQHAANVEKLLSLPEYVYPTIHLANDPYYKVODIEQLKDYKEC 1020
DB 961 CLVKNINVRREYLKQHAANVEKLLSLPEYVYPTIHLANDPYYKVODIEQLKDYKEC 1020
OY 1021 LMFVLEILMAKNNNSHAFIRKRVENIKQTKDAQGPDDAKMNKLTVCDDVANNITMS 1080
DB 1021 LMFVLEILMAKNNNSHAFIRKRVENIKQTKDAQGPDDAKMNKLTVCDDVANNITMS 1080
OY 1081 TTYSLSPKDPVLPAPFTQPDKNFNTKYLRLPEPKSFPTPKPTTNVLGAVNPLSS 1140
DB 1081 TTYSLSPKDPVLPAPFTQPDKNFNTKYLRLPEPKSFPTPKPTTNVLGAVNPLSS 1140
OY 1141 AGKOSQTKSSRMETVSNASSSSNPSPGRIKGLDSEMDHSENEDEYTWSSPLPGKSSDK 1200
DB 1141 AGKOSQTKSSRMETVSNASSSSNPSPGRIKGLDSEMDHSENEDEYTWSSPLPGKSSDK 1200
OY 1201 RDDSOLVRSLEKPRGRKTRPYTEQBEKLGMDLTLYLQOKKGSQSRSKRGHTASESD 1260
DB 1201 RDDSOLVRSLEKPRGRKTRPYTEQBEKLGMDLTLYLQOKKGSQSRSKRGHTASESD 1260
OY 1261 EOOMPEEKRLKEDILENEDBONSPPKKGKRGPRPKPLGGGTPEKEPTMTSKKSGSKKSG 1320
DB 1261 EOOMPEEKRLKEDILENEDBONSPPKKGKRGPRPKPLGGGTPEKEPTMTSKKSGSKKSG 1320


```
QY 1321 PPAPEEEEROSGNTGOKSKOHRVSRRAQRAESPESAIESTOSTPGKGRPSK 1380
    |||||||
Db 1321 PPAPEEEEROSGNTGOKSKOHRVSRRAQRAESPESAIESTOSTPGKGRPSK 1380
QY 1381 TPSPSOPKKNV 1391
    |||||||
Db 1381 TPSPSOPKKNV 1391

RESULT 2
ABG72803
ID ABG72803 standard; Protein: 1297 AA.
AC ABG72803;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human tumour suppressor SCC-112.
XX
KW Human: SCC-112; cancer: apoptosis; alldodynia; cell proliferation;
KW degenerative disorder; metastasis inhibition; breast cancer; causalgia;
KW kidney cancer; bladder cancer; pancreatic cancer; colon cancer;
KW squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;
KW global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia;
KW hypoxia-induced nerve cell damage; anxiety; diabetes mellitus;
KW cardiac arrest; spinal cord lesion; lung cancer; tumour suppressor;
KW neonatal distress; Alzheimer's disease; colon cancer; stomach cancer;
KW multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;
KW Huntington's disease; Parkinson's disease; Korsakoff's syndrome;
KW amyotrophic lateral sclerosis; cell survival; uterine cancer.
XX
OS Homo sapiens.
XX
FH Key
FH Location/Qualifiers
FT 2..137
FT /Label= RhogEF domain
FT 127..651
FT /Label= N-adaplin_domain
FT 166..187
FT /Label= Leucine_2ipper
FT 249..350
FT /Label= SKP1_domain
FT 597..617
FT /note= "PEST sequence site"
FT 858..865
FT /note= "Tyrosine kinase phosphorylation site"
FT 920..926
FT /note= "Nuclear localisation signature sequence"
FT 1030..1036
FT /note= "Tyrosine kinase phosphorylation site"
FT 1143..1663
FT /note= "PEST sequence site"
FT 1216..1277
FT /note= "PEST sequence site"
FT 1225..1231
FT /note= "Nuclear localisation signature sequence"
FT 1227..1230
FT /note= "Nuclear localisation signature sequence"
FT 1228..1234
FT /note= "Nuclear localisation signature sequence"
FT 1232..1235
FT /note= "Nuclear localisation signature sequence"
FT 1251..1257
FT /note= "Nuclear localisation signature sequence"
FT 1278..1297
FT /Label= Epitope_region
FT /note= "Residues 1278 to 1297 are
specifically claimed in claim 41"
XX
PN WO200281641-A2.
XX
PD 17-OCT-2002.
XX
```

```
PF 08-APR-2002; 2002WO-US10850.
XX
XX 06-APR-2001; 2001US-281780P.
PR
XX (GEOU ) UNIV GEORGETOWN.
XX
XX Kasid UN, Kumar D, Ahmad I;
XX
XX WPI: 2003-103330/09.
DR
DR N-PSDB: ABX14056.
XX
PT New isolated SSC (undefined) tumor suppressor polypeptides and
PT polynucleotides, useful for diagnosing, preventing or treating cancer
PT or degenerative disease, e.g. Alzheimer's Disease, Huntington's
PT disease, or multiple sclerosis
XX
XX Claim 13; Fig 1B; 83pp; English.
XX
XX
```

The invention relates to a new isolated polypeptide SCC-112. The SCC-112 polypeptides and polynucleotides are useful for diagnosing, preventing or treating cancer (e.g. breast cancer, kidney cancer, bladder cancer, pancreatic cancer, colon cancer, squamous cell carcinoma, uterine cancer, stomach cancer, colon cancer, lung cancer); or degenerative disease or disorder (e.g. global and focal ischaemic and haemorrhagic stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell damage, nerve cell damage caused by cardiac arrest or neonatal distress, epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb pain, causalgia, neuralgias, herpes zoster, spinal cord lesions, hyperalgesia, alldodynia, Alzheimer's disease, Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and Korsakoff's syndrome). The polypeptides and polynucleotides are also useful for inducing apoptosis in cancer cells, increasing survival or proliferation of a cell, or inhibiting cancer cell proliferation and/or metastasis in a cancer patient. The polynucleotides can be used as probes to detect complementary nucleotide sequences, or as primers to obtain additional copies of the polynucleotides. SCC-112 may also be used for identifying drugs for treatment of cancers. The present sequence represents the amino acid sequence of human tumour suppressor SCC-112.

Sequence 1297 AA;

Query Match 64.5%; Score 4642.5; DB 24; Length 1297;

Best local similarity 68.8%; Pred. No. 3.7e-299;

Matches 905; Conservative 159; Mismatches 203; Indels 49; Gaps 11;

```
QY 31 MYRLKVVYKTPMDMDSEEEKLYLNLAHLASDFLKHGKVRLVACCLADIFRI 90
    :::::::::::::::
Db 1 MIKRLKVVYKTPMDMDSEDEKOOYLLPLALHLASEFFLRNPKDVRLLVACCLADIFRI 60
QY 91 YAPEAPYSPDKLKDIPMTTPROLKGLGDEFTKSPQNRVFFYLENLAWKSVICPELEDS 150
    |||||||
Db 61 YAPEAPYSHDKLKDIPMTTPROLKGLGDEFTKSPQNRVFFYLENLAWKSVICPELEDC 120
QY 151 NEIFQRLTFLFSVYINNGHGNOKVHMHVMDLMSIICEDGTYSQELLDVYLVVPAHKNL 210
    |||
Db 121 NEIFQRLTFLFSVYINNGHGNKVKVGMHMDLMSIIMEGQVYQELLSLNLIRAHKNL 180
QY 211 NKQAYDLAKALIKRTQAQIEPYITTFPNOVLMGTSISDUSEHVFDLILELYNIDSLL 270
    |||
Db 181 NKQSPDLAKVLLKRTVQRTIACIANFNOVYLVGRSSVSDISEGVFDILOEFAIDPHLL 240
QY 271 LSVIPOLPEFKLSNNEERLOVYKLLAMFPGKDSLELSONKPLMOCYVIGRNDIHVPR 330
    |||
Db 241 LSVMPOLPEFKLSNGEERLAVYKLLAVKSDSLATONKPLMOCYVIGRNDIHVPR 300
QY 331 LECVFASHCLMNHDLAKDLTEYLKVRSHDPEEAIRHDVYSIYTAARKDILLVNDHL 390
    |||
Db 301 LESVAFASHCLMNHDLAKDLTEYLKVRSHDPEEAIRHDVYITITAAKRDALVNDQL 360
QY 391 NFVRERTLDKRRVRKEMAGLQYKYKYLQSAAGKDAQOIAIKRLHIYQNSID 450
    |||
Db 361 GFVRERTLDKRRVRKEMAGLQYKYKYLHGEAGKEBAEKVSWIKDLHIYQNSID 420
```

```

QY 451 DELLVERIFAQYVWPHNLLETTERMKCLYLATLDLNAVKALNEMKCONLLRHQVKDL 510
Db 421 DLLVEKIFAQYLVPHNLLETTERMKCLYLATSLDPNAVKALNEMKCONLLRSHVRELL 480
QY 511 DLKOPRTDASVKAIFSKVAVITRNLPDPGKADQFMKFTQVLEDEKIRKOLEVLYSPT 570
Db 481 DLKOPRTSEANCSAMFGLMTIAKNLPDPGKADQFMKFTQVLEDEKIRKOLEVLYSPT 540
QY 571 CSCQKQACGCVREITTKRKNPKNPFLNEMIKFLLEETIAVPHIDTESIALIKOVNSID 630
Db 541 CSCQKQADICVREIARLANRKNPFLNEMIKFLLEETIAVPHIDTESIALIKOVNSID 600
QY 631 GTADDEEGVPTDQAIRAGLELLKVLSTFHPISFSAETESLACLKMDKDEKVAEALQ 690
Db 601 GTADDEEGVSPTRAIRSGLELLKVLSTFHPISFSAETESLACLKMDKDEKVAEALQ 660
QY 691 IFKNTGSKIEEDPHITSALLPVILHKSCKGPPQAKYAIHCHAITSSKETQAPQIFEP 750
Db 661 IFKNTGSKIEEDPHITSALLPVILHKSCKGPPQAKYAIHCHAITSSKETQAPQIFEP 720
QY 751 LKHSIDPSNLEHLLTPTVIGITALLPDQFAAPKSMVATFIVKDLMDRLPGKTKTK 810
Db 721 LKHSIDPSNLEHLLTPTVIGITALLPDQFAAPKSMVATFIVKDLMDRLPGKTKTK 780
QY 811 LWPDEEVSPEPTMYKQAIKMMVWMLGMKNNSKSGSTFLRLTLTTHSDGLTTEQGI 870
Db 781 LWPDEEVSPEPTMYKQAIKMMVWMLGMKNNSKSGSTFLRLTLTTHSDGLTTEQGI 840
QY 871 SKPDMSTRILAGSAIYKLAOEPCYHEITLEQOICALAINECYQOVQVFAQKLEKGL 930
Db 841 SKPDMSTRILAGSAIYKLAOEPCYHEITLEQOICALAINECYQOVQVFAQKLEKGL 900
QY 931 SRLRLPLEYNAICALAKDPVKERRAHAROCLVNINVRREYLKQHAASVEKLSLPEY 990
Db 901 VKLLPLEYNAIFALCKDPVKERRAHAROCLVNINVRREYLKQHAASVEKLSLPEY 960
QY 991 VPPYTHLLAHDPYKAVODIEOLKDYECLEMLFVETLMAKNENNSHAFIRKVENIKQT 1050
Db 961 VPPYTHLLAHDPYKAVODIEOLKDYECLEMLFVETLMAKNENNSHAFIRKVENIKQT 1020
QY 1051 KDNQGPDAKMEKLYVCVAMNIIKSKSTTYSLESKPKDVLPAPEFQDPKKNSTKN 1110
Db 1021 KDNQGPDAKMEKLYVCVAMNIIKSKSTTYSLESKPKDVLPAPEFQDPKKNSTKN 1080
QY 1111 YLPEPMKSFPTPKPKTTNVLGAVNPKLSAGKOSQTKSRMETVS--NASSSSNPSPPG 1168
Db 1081 YLPEPMKSFPTPKPKTTNVLGAVNPKLSAGKOSQTKSRMETVS--NASSSSNPSPPG 1140
QY 1169 RIKGR-LDSSEMHSENE--YTMSSPLPKKSKRDSPLVNSELEKPRGKRTPYTEQ 1225
Db 1141 RIKGR-LDSSEMHSENE--YTMSSPLPKKSKRDSPLVNSELEKPRGKRTPYTEQ 1200
QY 1225 EERKGMODLKLVOEQPKGORSRKRGHVASEDE--QOMPEERKLKEDILNEDENSP 1284
Db 1180 EERKGMODLKLVOEQPKGORSRKRGHVASEDE--QOMPEERKLKEDILNEDENSP 1260
QY 1285 -PKKGRGRPPKPLG--GTPKEEPTMKTSRKSGSKSGPPAPEEEREROSGNTG 1338
Db 1224 -PKKGRGRPPKPLG--GTPKEEPTMKTSRKSGSKSGPPAPEEEREROSGNTG 1275

```

```

RESULT 3
ABP64685
ID ABP64685 standard; Protein: 919 AA.
XX
XX ABP64685:
XX
XX 25-FEB-2003 (first entry)
XX
XX Human protein SEQ ID 345.
XX
XX Human: expressed sequence tag: EST;
KW
KW haematopoietic disorder; central nervous system disease; viral infection;

```

```

KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; nontopic;
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW cytosolic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective.
OS Homo sapiens.
PN WC0200259260-A2.
XX
XX 01-AUG-2002.
XX
XX 16-NOV-2001; 2001WO-0542950.
XX
XX 17-NOV-2000; 2000US-0714936.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QX;
XX Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-590824/63.
XX
XX N-PSDB: ABQ99271.
XX
XX New isolated polynucleotide, useful in research, diagnostic or
XX therapeutic methods, e.g. preventing or treating disorders involving
XX aberrant protein expression or biological activity -
XX
XX Claim 20: SEQ ID 345; 394pp; English.
XX
XX The present invention relates to novel human coding sequences
XX (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are
XX useful in therapeutic, diagnostic and research methods. The
XX polynucleotides may be used in the field of molecular biology as
XX hybridisation probes, primers for PCR, for chromosome and gene mapping,
XX for the recombinant production of protein, or in generation of anti-sense
XX DNA or RNA. The polynucleotides are useful in diagnostics as expressed
XX sequence tags (ESTs) for identifying expressed genes or for physical
XX mapping of the human genome. The proteins may be used as molecular weight
XX markers, or as nutritional sources or supplements. The proteins may be
XX used to maintain and expand cell population in a totipotent or
XX pluripotent state useful for re-engineering damaged or diseased
XX tissues, transplantation, manufacture of bio-pharmaceuticals or the
XX development of bio-sensors. The polynucleotides and proteins are useful
XX for preventing, treating or ameliorating disorders involving aberrant
XX protein expression or biological activity, e.g. haematopoietic disorders,
XX central/peripheral nervous system diseases, mechanical and traumatic
XX disorders, non-healing wounds, immune deficiencies and disorders,
XX infectious diseases caused by viral, bacterial or fungal infection,
XX autoimmune disorders, allergic reactions and conditions, coagulation
XX disorders, or cancer. The polynucleotide sequences of the invention were
XX assembled from ESTs isolated mainly by sequencing by hybridisation, and
XX in some cases, sequences obtained from one or more public databases.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 919 AA;

```

```

Query Match 41.7%; Score 2999.5; DB 23; Length 919;
Best Local Similarly 63.3%; Pred. No. 2,7e-190;
Matches 594; Conservative 123; Mismatches 172; Indels 49; Gaps 11;

```

```

QY 409 MMGLAOIYKRYALQSAAGKRAAKOIMIKRKLHIYYONSIDRLVERIFAQYVWPHNL 468
Db 1 MMGLAOIYKRYALQSAAGKRAAKOIMIKRKLHIYYONSIDRLVERIFAQYVWPHNL 60
QY 469 ETTERMKCLYLYATLDLNAVKALNEMKCONLLRHQVKDLDPILKOPKTDASVKAIFSK 528
Db 61 ETTERMKCLYLYATLDLNAVKALNEMKCONLLRHQVKDLDPILKOPKTDASVKAIFSK 120
QY 529 VAVITRNLPDPGKADQFMKFTQVLEDEKIRKOLEVLYSPTSCQKQACGCVREITTKLG 588

```

Db	121	LMTIAKNLPDEKQADPVKKNQVLCDDDEKRLSQLELLISPCSCQADICVREIRAKRLA	180	KW	actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
Qy	589	NKQPTNPFLMIKIKLLERLARIYVHDTEISALIKOVNKSIDGTADDEEGVPTDOAIRA	648	KW	clitrorrhosis; hepatitis; mixed connective tissue disease; MCTD;
Db	181	NKQPTNPFLMIKIKLLERLARIYVHDTEISALIKOVNKSIDGTADDEEGVPTDOAIRA	240	KW	myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
Qy	649	GLELLKVSFTPHISPHSAETFEESLLACMKMDEKVAEALQIFKTKSGKIEDPHIRS	708	KW	autoimmune disease; inflammation; acquired immunodeficiency syndrome;
Db	241	GLELLKVSFTPHISPHSAETFEESLLACMKMDEKVAEALQIFKTKSGKIEDPHIRS	300	KW	AIDS; Addison's disease; adult respiratory distress syndrome;
Qy	709	ALLPVLHHSKKGPPROAKYAIHCHIAIFSSKETQFAQTEPDLHSLDPSNLEHLLTPV	768	KW	allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
Db	301	TLIPILHOKAKRGTHQAKQAVHCHIAIFTKNEVQAOQFEELSRSLANDVEEQLITPLV	360	KW	Werner syndrome, trauma; human.
Qy	769	TIGHIALAPDOFAPAPKWSVATFYKDLMDRLPGKTKTKLWVPDEEVSEPTMYKIOA	828	OS	Homo sapiens.
Db	361	SLGHSMLAPDOFAPAPKWSVATFYKDLMDRLPGKTKTKLWVPDEEVSEPTMYKIOA	420	XX	
Qy	829	IKMAYRWLLGMKNNSKSGTSLRLTLTTLHSDGLTEGKTSKPDMSRLAAGSALYK	888	XX	
Db	421	IKLTVRWLLGMKNNSKSGTSLRLTLTTLHSDGLTEGKTSKPDMSRLAAGSALYK	480	XX	
Qy	889	LAQEPCHYEITLPEYOICALALAINDECYQVRFQAKLHGLSRLPLPEYMAICALCAK	948	FT	Key
Db	481	LAQEPCHYEITLPEYOICALALAINDECYQVRFQAKLHGLSRLPLPEYMAICALCAK	540	FT	Location/Qualifiers
Qy	949	DPVKERRAAROCCLVKNINVRREYIKOHANASEKLLSLPERVVPYTHLADHPDYKV	1008	FT	48
Db	541	DPVKERRAAROCCLVKNINVRREYIKOHANASEKLLSLPERVVPYTHLADHPDYKV	600	FT	/note= "Potential glycosylation site"
Qy	1009	ODIEOLKDVKECLMFVLETLMAKNNSHAFIRKRVENIKOTKAOGPDADAKNEKLYTV	1068	FT	88
Db	601	ODVODLRIKECLMPLMELVMTKNNSHAFIRKRVENIKOTKAOGPDADAKNEKLYTV	660	FT	/note= "Potential phosphorylation site"
Qy	1069	CDVANNIIMSKSTYSLSPPKDPVLPAREFTOPDKNFSNTKNYLPPEMKSFTPGKPKTT	1128	FT	99
Db	661	CDVALCVINSKALCNADSPKDPVLPAREFTOPDKNFSNTKNYLPPEMKSFTPGKPKTT	720	FT	/note= "Potential phosphorylation site"
Qy	1129	NVLGVANRPPLSAGQOSQTSKSRMETVS--NASSSSNPSSGRIKGR-LDSEMDHSENE	1185	FT	153
Db	721	NVLGVANRPPLSAGQOSQTSKSRMETVS--NASSSSNPSSGRIKGR-LDSEMDHSENE	780	FT	159
Qy	1186	D--YTMSSPLPGKSKDKRDSDLVRSLEKPRGRKTPVTEQEKLGMDLTKLVQEQKP	1243	FT	/note= "Potential phosphorylation site"
Db	781	ENPVRIISVPEVKNTD-----PYKNKE--INSDOAT-----QGN	812	FT	422
Qy	1244	KGSORRRKRGHTASEDE--QOWPEEKRULEKEDILENEDQNSP--PKKGRGRPKPLG--CG	1300	FT	/note= "Potential phosphorylation site"
Db	813	ISSORGRKRTVTAAGAENIQOKTDEK-----VDESGPPAPSKPRGRPKSESQCN	863	FT	492
Qy	1301	TPKEPTMKTSKSGSKSGPPAPEEEROSGNTG	1338	FT	/note= "Potential phosphorylation site"
Db	864	ATKNDLMLPKPKGRKR---AAVGQESPPGLEAGNAK	897	FT	/note= "Potential glycosylation site"
RESULT 4					
AAB01382					
AAB01382 standard; Protein: 854 AA.					
AAB01382;					
20-OCT-2000 (first entry)					
Neuron-associated protein.					
Neuron associated protein; NEUAP; neurological disorder; epilepsy;					
ischemic cerebrovascular disease; stroke; cerebral neoplasm;					
Alzheimer's disease; Pick's disease; Huntington's disease;					
dementia; Parkinson's disease; demyelinating disease; meningitis;					
prion disease; Kuru; Creutzfeldt-Jakob disease; neurofibromatosis;					
cerebral palsy; muscular dystrophy; central nervous system; CNS;					
peripheral nervous system; PNS; myopathy; schizophrenia;					

FT Modified-site 697 /note= "Potential glycosylation site"
FT Modified-site 712 /note= "Potential phosphorylation site"
FT Modified-site 747 /note= "Potential glycosylation site"
FT Modified-site 750 /note= "Potential phosphorylation site"
FT Modified-site 757 /note= "Potential phosphorylation site"
FT Modified-site 798 /note= "Potential phosphorylation site"
FT Modified-site 800 /note= "Potential glycosylation site"
FT Modified-site /note= "Potential phosphorylation site"
XX MO20003447-A2.
XX 15-JUN-2000.
XX 10-DEC-1999; 99WO-US0408.
XX 11-DEC-1998; 98US-0210083.
PR 11-DEC-1998; 98US-9123456.
PR 09-FEB-1999; 99US-0119365.
PR 16-MAR-1999; 99US-0124687.
XX (INCY-) INCYTE PHARM INC.
XX Tang YT, Xue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
PI Lu DM, Azimzal Y;
XX WPI; 2000-423423/36.
DR N-PSDB; AAA47423.
XX New human neuron-associated proteins and polynucleotides encoding them,
PT useful for diagnosis, treatment and prevention of cell proliferative
PT disorders including cancer, neuronal and neurological disorders
XX Claim 1; Page 103-105; 145pp; English.
XX Human neuron-associated proteins (NEUAP) can be used for for
CC treating or preventing a disorder associated with decreased
CC expression or activity of NEUAP. Antagonists of NEUAP are useful for
CC treating or preventing disorder associated with increased expression
CC or activity of NEUAP. NEUAP or their fragments or derivatives are
CC useful for treating neurological disorder such as epilepsy, ischemic
CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC disease, Pick's disease, Huntington's disease, dementia and
CC Parkinson's disease. NEUAPs are also useful for treating other
CC demyelinating diseases, bacterial and viral meningitis, prion
CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC metabolic diseases of the nervous system, neurofibromatosis, other
CC developmental disorders of the central nervous system, cerebral
CC palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC other neuromuscular disorders, peripheral nervous system disorders,
CC inherited, metabolic, endocrine, and toxic myopathies, mental
CC disorders including mood, anxiety and schizophrenic disorders, a cell
CC proliferative disorder such as actinic keratosis, arteriosclerosis,
CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC hemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC disorder such as acquired immunodeficiency syndrome (AIDS). Addison's
CC disease, adult respiratory distress syndrome, allergies, ankylosing
CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC complications of cancer, hemodialysis, and extracorporeal circulation,
CC viral, bacterial, fungal, parasitic, protozoal, and helminthic
CC infections, and trauma. This protein was given the Incyte ID no.
XX 2888437CD1.
XX Sequence 854 AA;

Query Match 37.8%; Score 2715.5; DB 21; Length 854;
Best Local Similarity 62.2%; Pred. No. 1.7e-171;
Matches 543; Conservative 114; Mismatches 167; Indels 49; Gaps 11;
QY 474 MKCIYIYATLIDNAVALNEMKCOULLRHQVKDLDLIDKOPTDASVAIRSKVAVIT 533
DB 1 MKCIYLYASLIDPAVVALNEMKCOUMLRHVELLDLHKOPTSEANCSAMFGKMLTIA 60
QY 534 RNLPPGKADPFMKKFVLEDEDEKIRKQLEVLVSPYSCGQAGCVREITKKLGANKOP 593
DB 61 KNLPPGKADPFVKFQVGLDDEDEKLRQLELLSPYSCGQADICREIARKLANKOP 120
QY 594 TNPLEMIKELLERIAVPHIDTESISALIKOVNKSIDGTADDEDEGVPTDQAIRAGLELL 653
DB 121 TNPLEMVKFLERIAVPHIDSEASIALVLMKNSIGTADDEEGVSPDIAIRSGLELL 180
QY 654 KVLSTPIPISEHSETESLCLAKMDEKVAEALDIFKNTGSKIEDDEPHIRALLPV 713
DB 181 KVLSTPIPISEHSETESLCLAKMDEKVAEALDIFKNTGSKIEDDEPHIRALLPV 240
QY 714 LHHKSKGPPROAKYAIHCHAIRPSKEQPAQFEEDLHKSILDSNLEHITPLVTGHI 773
DB 241 LHHKSKGPPROAKYAIHCHAIRPSKEQPAQFEEDLHKSILDSNLEHITPLVTGHI 300
QY 774 ALAPDQFAAPKWSVATFTVKDLMMDRLPGKKTTRKLVWPDEEVSPETVYKIOAIKMY 833
DB 301 SMLAPDQFAAPKWSVATFTVKDLMMDRLPGKKTTRKLVWPDEEVSPETVYKIOAIKMY 360
QY 834 RMLGKNNNSKSTSTRLTLTLHSDGDLTEBGKISKDMSRLKLAASATYKLAQEP 893
DB 361 RMLGKNNNSKSTSTSTRLTLTLHSDGDLTEBGKISKDMSRLKLAASATYKLAQEP 420
QY 894 CYHEITITPEOFCALAINDECVYOVROVPFOKLHKGISRLPLEYATGALCKADPYKE 953
DB 421 CYHEITITPEOFCALAINDECVYOVROVPFOKLHKGISRLPLEYATGALCKADPYKE 480
QY 954 RRAHAROCVKNINVRREYLKQHAASEKLLSLPEVYVPTTHLADHPYKVQDIEQ 1013
DB 481 RRAHAROCVKNINVRREYLKQHAASEKLLSLPEVYVPTTHLADHPYKVQDIEQ 540
QY 1014 LKDYKECLMIVYLELMAKNENSHAFTRKVENIKOTKRDQOPDAMNEKLTIVCYAM 1073
DB 541 LKDYKECLMIVYLELMAKNENSHAFTRKVENIKOTKRDQOPDAMNEKLTIVCYAM 600
QY 1074 NIIMSKSTYSLESPKQVPLPARFPTQPDKNFSWTKVYLPLPEMSFTTPGKPTTVIGA 1133
DB 601 NIIMSKSTYSLESPKQVPLPARFPTQPDKNFSWTKVYLPLPEMSFTTPGKPTTVIGA 660
QY 1134 VNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENE--YT 1188
DB 661 VNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENE--YT 1248
QY 1189 MSSFLPKKSDKRDSDSLVSELEKPRGKKTPTTBEDEKLGMDLTKLVQEQPKRSOR 1248
DB 721 MSSFLPKKSDKRDSDSLVSELEKPRGKKTPTTBEDEKLGMDLTKLVQEQPKRSOR 752
QY 1249 SRKKGHVASDSE-QQWPEEKRLKEDILENDEBONS-PKGGKGRPPKPLG-CGTPEE 1305
DB 753 SRKKGHVASDSE-QQWPEEKRLKEDILENDEBONS-PKGGKGRPPKPLG-CGTPEE 803
QY 1306 PYMTSKSKSKKSGPPAPEEEEROSGNTG 1338
DB 804 PYMTSKSKSKKSGPPAPEEEEROSGNTG 832
RESULT 5
ABB67869
ID ABB67869 standard; Protein: 1218 AA.
XX ABB67869;
AC ABB67869;
XX 26-MAR-2002 (first entry)
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 30399.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 XX
 DR N-PSDB; ABL11972.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure: SEQ ID NO 30399; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL161/6-ABL30511), expressed DNA
 CC sequences (AB57737-AB572072).
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 1218 AA;
 SQ
 Query Match 27.6%; Score 1982.5; DB 22; Length 1218;
 Best Local Similarity 37.3%; Pred. No. 1,3e-122;
 Matches 444; Conservative 247; Mismatches 421; Indels 77; Gaps 22;

DB 362 VMAIVETAKRDFTVLEAPDLEIVRETRLDKKYKIRDMANGLAIVYKRAICEPNDIST 421
 QY 429 AAK-QIAWIKDKLHIYTONSIDRLYERIFPAQYVPHNLETTRMKCLIVYATLDLN 487
 DB 422 GLKRVSDWIKNNKILGYSKVGLIEDRLVERLLDITGLVPKLAPBERMKLYHLGDLDDAN 481
 QY 488 AVKALNEMKKCONLRLHGVKDLDDIKQPK-TDASVKAFISVWVITRNLDPGKAQDPM 546
 DB 482 ATKAEVELOKQMKTRNVSDWIKLHHSKEPTPRVLSOISAQANIATKLDPDLAAEYL 541
 QY 547 KKFQVLEDEDEKIRKQLEVLVSPSCQAQECVREITKLGPNKOPTNPELEMTKFFLE 606
 DB 542 TQFSNNLRKRDQGLKCNINVLKRDVSCHECADPMGLLKGAAHQ-SLUVNYTKMLIE 600
 QY 607 RIAPVHIDTESIALIKOVNKSIDGTADDEDCVPTDQAIRAGLELLKVLSTPHRISHS 666
 DB 601 RVASVWVVKESIGVILSIEQIEKSGMCEEIGISAQEGERGLKMLSVFSANHFPT 660
 QY 667 AEFESLLACLMDEKVAEALOLFKNMG---SKIEDFPPIRSALLPVLHKSKKGP 723
 DB 661 DTSLRHLISLSYEDDYVAPLVKTLTHGKQPLIDPTPAIDELAPVCKDFALITGP 720
 QY 724 RQAKYAIIHCIH-----AIFSKETQ-----FAQIFEPHLKSLDPSNLEHL 763
 DB 721 KOAKHAVKCIFFVNSQSASTDQATSGAGSASTPTQVPHIFENIETRLKLTLP-NCEHQ 779
 QY 764 ITPLVITIGHIALAPDOFAAPKRSVAFIYKDLMDRLGKTKTKL-----WPDDEV 818
 DB 780 RTKIYTLGHIAVYVMPQAFELTPKNNIARIYKELLIOE-VPAQORYELPEDSDMCAQERL 838
 QY 819 SPEWVKIQAIKMWVYKNNKNSKSTSLRLTLTILSHSDGLTEQGISKPDMSRL 878
 DB 839 PPDTICKIDALKAAARWMLGLRTD-EHAAQKTFRLAFAVWORGDLLONRICGAKESWL 897
 QY 879 RLAAQSATVLAQEDCYHEITLLEQYOLCALAINDECYQVQVFAQKLNKLSRLR---- 934
 DB 898 RLGAACAMLKVCQGVGSDQYSAEQYLQLSQADMAPVEVEIRFARKLHKLSRLPRNC 957
 QY 935 LPLEVMALCALCAPVVERRAHARQCLVKNINVARRELYKQHAANS-----EKLSTLP 988
 DB 958 LPDFMGVLVYLAGLETERKRLDVLRYHVAETVYVNNKRELYKTVAMTSPDSSTESQSLHLP 1017
 QY 989 EYVVPYTIHLHADDDYKVDIDQKQVCKECLMVLIELMAKNNSHNSHAFKRVENIK 1048
 DB 1018 DYMLAFAIPVLVHPDRFNNHEDYVQLRMEKCLRTLEPLAKRTTFYHSYKQLOQLIK 1077
 QY 1049 QTKDAQGPDDAKMKNEKLYTCVDVANNIISK-----STYVLESPPDVLPAHFPTQ 1101
 DB 1078 HREFSLG-SDKRDVYKMAALCDLAMYIIDSKFSPPDGTSTFSPML-----ALPEMYKEP 1132
 QY 1102 D-KNPSNKNLUPPEMKSFPTPGKPKTTNVGVAVKPLSSAKOSQTS 1149
 DB 1133 AVANFQNDVYIPLDVT-----LGA--KSTSKAAATAMTTS 1167
 RESULT 6
 AAU27827
 ID AAU27827 standard; Protein: 217 AA.
 AC AAU27827;
 DT 18-DEC-2001 (first entry)
 DE Human full-length polypeptide sequence #152.
 XX
 XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
 KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
 KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
 KW nervous system disorder; inflammatory disorder; cell differentiation;
 KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
 KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
 KW cytosolic; antihemmatic; antiarthritic; vulnerability; inflammatory;
 KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;

KW neuroprotective; osteopathic; antidiabetic; antisthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX Homo sapiens.
PN WO200164834-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04926.
XX
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAR-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Dmanac R;
XX
XX WPI; 2001-569862/66.
DR N-PSDB; AAS44727.
XX
XX Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
XX Claim 10; SEQ ID NO 324; 153pp; English.
XX
XX Sequences AAU27676-AU28019 represent full-length polypeptides and
CC contig polypeptides of the invention. The proteins and their associated
CC DNA sequences are useful for the treatment, diagnosis and prevention of
CC various types of disorder in a mammalian subject such as a human, dog,
CC monkey, mouse, hamster or rat. The disorders include cancers such as
CC leukemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as hepatitis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 217 AA;
Query Match 11.7%; Score 839; DB 22; Length 217;
Best Local Similarity 100.0%; Pred. No. 9.7e-48;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
AAB58462
ID AAB58462 standard; Protein: 363 AA.
XX
XX AAB58462;
XX
XX 14-MAR-2001 (first entry)
XX
XX
DE Lung cancer associated polypeptide sequence SEQ ID 800.
XX
XX Human; lung cancer associated protein; neuroprotective; cytoskeletal;
XX cardiocactive; immunomodulatory; muscular active; vulnereary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
XX Homo sapiens.
XX
XX WO20005180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
DR N-PSDB; AAF18338.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX
PS Claim 11; Page 1337-1339; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardiocactive;
CC immunomodulatory; muscular active general; vulnereary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
XX
SQ Sequence 363 AA;
Query Match 10.5%; Score 758.5; DB 21; Length 363;
Best Local Similarity 45.0%; Pred. No. 4.3e-42;
Matches 172; Conservative 56; Mismatches 105; Indels 49; Gaps 11;

QY 1231 MDDLTKLVQOKPKGSGRSRRKRGHTASESDQOMPEEKRLKEDILENEDQNSPPKKGKR 1290
DB 1 MDDLTKLVQOKPKGSGRSRRKRGHTASESDQOMPEEKRLKEDILENEDQNSPPKKGKR 60
QY 1291 GRPPKPLGGGTPEEPEPMKTSKKGSKKSGPPAPEEEEBEEROSGNTQKSKSKQHVSR 1350
DB 61 GRPPKPLGGGTPEEPEPMKTSKKGSKKSGPPAPEEEEBEEROSGNTQKSKSKQHVSR 120
QY 1351 RAOORAEPSSESAIESTQTPPOKGRGRSPKTPSPSOQKKNV 1391
DB 121 RAOORAEPSSESAIESTQTPPOKGRGRSPKTPSPSOQKKNV 161

QY 965 NINWREYLYKQHAHVSEKLLSLPEYVVPPTIHLAHDPYVYVODIEQKDKVCEKLMFV 1024
DB 1 NISIRREYIKQNMWATEKLLSLPEYVVPPTIHLAHDPYVYVODIEQKDKVCEKLMFV 60
QY 1025 LETIMAKENNSHAFLIKWVENIKQTKDAQCPDDAKNEKLYVYVADYAMNIIKSKSTTVS 1084

[illegible]

```

RESULT 8
ABB84560
ID ABB84560 standard; Protein: 182 AA.
XX
AC ABB84560;
XX
DT 23-JAN-2003 (first entry)
DE Human euchromosome fragile intelligence delay syndrome protein 20.02.
XX
KW Euchromosome fragile intelligence delay syndrome protein 20.02; human;
KM euchromosome fragile intelligence delay syndrome; dementia.
XX
OS Homo sapiens.
PN CNI351041-A.
PD 29-MAY-2002.
XX
PF 26-OCT-2000; 2000CN-0125797.
XX
PR 26-OCT-2000; 2000CN-0125797.
XX
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
PI Mao Y, Xie Y;
XX
DR WPI: 2002-637117/69.
DR N-PADB; ABS56684.
XX
PT Human euchromosome fragile intelligence delay syndrome associated
PT protein 20.02 polypeptide, used to treat e.g. dementia -
XX
PS Claim 1; Page 30 (disclosure); 33pp; Chinese.
XX
CC This invention describes the novel human euchromosome fragile
CC intelligence delay syndrome associated protein, 20.02. The polypeptide is
CC used in treating diseases such as euchromosome fragile intelligence delay
CC syndrome, and dementia. This sequence represents the human euchromosome
CC fragile intelligence delay syndrome-associated protein, 20.02 described
CC in the disclosure of the invention.
SQ Sequence 182 AA;
Query Match 9.9%; Score 709; DB 23; Length 182;
Best Local Similarity 75.0%; Pred. No. 3.2e-39;
Matches 129; Conservative 24; Mismatches 19; Indels 0; Gaps 0;
409 MGCLAQIYKKRYALGSAAGKDAKOIAMIKDKLLHYQNSIDRLVERFQAQYVPHNL 468
|||||:|||||T |||::||::|||::|||::|||::|||::|||::|||::|||::|||

```

Dd	I	MGAIAQLYKRYCLHGEAGCKPAEAKVSVTKDKLLHIYYONSIDDKLLVEKIIFAQYLVPANL	60
Oy	469	EETPRMCIYLYVATTDLANAVKALNEMMKCONILRHQVDLDLDLIKOPTDASVAIAFSK	52
Dd	61	EETPRMCIYLYVASLDPNAVVKALNEMMKCONMLRSHVRRLDLHRQPISSEANCAMFCK	12
Oy	529	VMTITRNLPDGKAQDFMKKFPTOVLDEDERIKRQLEVLVSPTSCSKQAEGCV	580
		: :	
Dd	121	LMTIARKNLPDGKAQDFVKRFNOVLGDDEKLRSLQRELTISPSCSKQADICV	172
 RESULT 9 AAB94946			
ID	AAB94946	standard; Protein: 333 AA.	
AC	AAB94946;		
XX			
Dt	26-JUN-2001	(first entry)	
XX			
DE		Human protein sequence SEQ ID NO:16468.	
XX			
KW		Human; primer; detection; diagnosis; antisense therapy; gene therapy.	
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		
PD	07-FEB-2001.		
XX			
Pf	28-JUL-2000; 2000EP-0116126.		
XX			
PR	29-JUL-1999; 99JP-0248036.		
XX			
PR	27-AUG-1999; 99JP-0300253.		
XX			
PR	11-JAN-2000; 2000JP-0118776.		
XX			
PR	02-MAY-2000; 2000JP-0183767.		
XX			
PR	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishti S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
DR	WPJ: 2001-318749/34.		
XX			
Pt	Primer sets for synthesizing polynucleotides, particularly the 5602		
Pt	full-length cDNAs defined in the specification, and for the detection		
Pt	and/or diagnosis of the abnormality of the proteins encoded by the		
Pt	full-length cDNAs -		
PS	Claim 8; SEQ ID 16468; 2537pp + CD ROM; English.		
XX			
CC	The present invention describes primer sets for synthesising 5602		
CC	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dn primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	the specification. The primer sets can be used in antisense therapy and		
CC	in gene therapy. The primers are useful for synthesising polynucleotides		
CC	particularly full-length cDNAs. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13633 to AAH18742 represent human cDNA sequences: AAB92446 to		
CC	AAH95983 represent human amino acid sequences; and AAH13629 to AAH13632		
CC	represent oligonucleotides, all of which are used in the exemplification		
CC	of the present invention.		


```

XX SQ Sequence 333 AA;
Query Match
Best Local Similarity 42.7%; Pred. No. 2.6e-34;
Matches 150; Conservative 51; Mismatches 101; Indels 49; Gaps 11;

QY 996 IHLADPDVVKVQDIEQLKDKVECLFVLEILMAKNENNSHAFIRKRVENIKOTKDAQG 1055
Db 2 IHLADPDVTRSDQVDQLDLDICEJLMEVLEMTNENNSHAFIRKRVENIKOTKDAQG 61

QY 1056 PDDAKNNEKLYTCVDAAMNIMSKSTYTSLESFDPVLPARFPTQDPKNSNTKNVLPPE 1115
Db 62 PDSEKNEKLYTCVDAVALCYNKSKALCNADSPKDPVLPKPFPTQPEKDCNCKSYTSEE 121

QY 1116 MKSEFPGPKRTNVGAVNKPUSAGKOSQKTSRMEYTS--NASSSSMPSGRIRKGR 1173
Db 122 TRVLLITGKPKPGVIGAVNKPUSAGKOSQKTSRMEYTS--NASSSSMPSGRIRKGR 181

QY 1174 -LDSEMDHSENEED--YTMSSPLPGKSKSDKRDSDLVRSLEKPRGKRTPVTEOEKLG 1230
Db 182 SSEAETGVSENEENPVRIISVTPVKNID-----PVKNKE--IN 218

QY 1231 MDDLTLVDEQKRGSGORSKRGHTASEDE--QOMPEEKRLKEDILENEDEQNSP--PKKG 1288
Db 219 SDGAT-----QGNISSDRGKRRTVTAAGAEINIOQKTXDEK-----VDESGPAPASKP 264

QY 1289 KRGPRPKPLG--GGTPKEEPTMTKSKGSKKSGPPAPEEEDEREQSGNTE 1338
Db 265 RRGRRRKSESQGNATKNDLKRINKGRKRR-----AAVGQSPGGLPAGNAK 311

RESULT 10
AA001759
ID AA001759 standard; Protein: 165 AA.
XX AA001759;
XX 18-JUL-2001 (first entry)
XX Human secreted protein #38.
XX
XX Human: secreted protein; immunogen; antibody; diagnosis;
XX rheumatoid arthritis; hyperproliferative disorder; neoplasm;
XX cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
XX angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
XX fungal infection; corneal infection; wound healing; cell culture;
XX epithelial cell proliferation; skin ageing; transplantation;
XX tissue regeneration; chemotaxis; food additive.
XX
XX Homo sapiens.
XX
XX WO200123546-A1.
XX
XX 05-APR-2001.
XX
XX 26-SEP-2000; 2000WO-US26323.
XX
XX 27-SEP-1999; 99US-0155805.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben S, Komatsoulis GA;
XX
XX WPI: 2001-266150/27.
XX N-PSDB; AAS02396.
XX
XX Nucleic acids encoding 37 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX PS Disclosure; Page 454; 494pp; English.

```

```

CC The sequence represents a human secreted protein of the invention. The
CC polynucleotides, polypeptides and antibodies raised against them are used
CC to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polynucleotides and antibodies are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. The antibodies
CC can also be used in alleviating symptoms associated with the disorders
CC and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunoassay assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
XX
XX SQ Sequence 165 AA;
Query Match
Best Local Similarity 65.0%; Pred. No. 5.7e-32;
Matches 106; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

QY 964 KNINVRREYIKQAAASEKILSLPYVYPTTHLADPDVVKVQDIEQLKDKVECLWF 1023
Db 1 ENISIRREYIKQPMATEKLSLPEYVYPTTHLADPDVTRSDQVDQLRDKCECLWF 60

QY 1024 VLEILMAKNENNSHAFIRKRVENIKOTKDAQGPDPAKMEKLYTCVDAAMNIMSKSTY 1083
Db 61 MLEVLMTKNENNSHAFIRKRVENIKOTKDAQGPDPAKMEKLYTCVDAVALCYNKSKALC 120

QY 1084 SLESPKDPVLPARFPTQDPKNSNTKNVLPPEKKSFTTPGKPK 1126
Db 121 NADSPKDPVLPKMFPTQPEKDCNCKSYTSEETRVLLITGKPK 163

RESULT 11
ABB89260
ID ABB89260 standard; Protein: 147 AA.
XX ABB89260;
XX
XX 24-MAY-2002 (first entry)
XX
XX Human polypeptide SEQ ID NO 1636.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US16450.
XX
XX 19-MAY-2000; 2000US-205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX
XX WPI: 2002-122018/16.
XX N-PSDB; ABL89669.
XX

```

PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -

PS Claim 11: SEQ ID NO 1636; 2081bp + Sequence Listing; English.

XX The invention relates to novel genes (ABU9449-ABU90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 147 AA:

Query Match 8.2%; Score 591; DB 23; Length 147;
Best Local Similarity 96.6%; Pred. No. 1.6e-31;
Matches 115; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MAHSKRTNDGKITPPGVKEISDKISKEMVRLKMWVKFMDMDQSESEKELYTLA 60
Db 29 MAHSKRTNDGKITPPGVKEISDKISKEMVRLKMWVKFMDMDQSESEKELYTLA 88
Oy 61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYTPSPKLDKIDIFETIROLKGLD 119
Db 89 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYTPSPKLDKIDIFETIROLKGLD 147

RESULT 12

AA65247
ID AAY65247 standard; Protein: 101 AA.

XX AAY65247;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1408.

XX Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.

OS Homo sapiens.

PN M09953051-A2.

PD 21-OCT-1999.

PE 09-APR-1999; 99WO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-038446/03.

DR N-PSDB; AA242861.

XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
PT

PS Claim 3: Page 776; 837pp; English.

XX AA42265 to AA43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AA165438 represent the EST-related proteins corresponding to AA42265 to
CC AA43075. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis,
CC as well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA42249 to AA42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 101 AA:

Query Match 7.1%; Score 513; DB 21; Length 101;
Best Local Similarity 99.0%; Pred. No. 1.5e-26;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MAHSKRTNDGKITPPGVKEISDKISKEMVRLKMWVKFMDMDQSESEKELYTLA 60
Db 1 MAHSKRTNDGKITPPGVKEISDKISKEMVRLKMWVKFMDMDQSESEKELYTLA 60
Oy 61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYTPSP 100
Db 61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYTPSP 100

RESULT 13

AAG02811
ID AAG02811 standard; Protein: 92 AA.

XX AAG02811;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 6892.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PE 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

DR N-PSDB; AAC02817.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13: SEQ ID 6892; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
CC
SQ Sequence 92 AA;
OY Query Match 6.5%; Score 467; DB 21; Length 92;
Db Best Local Similarity 100.0%; Pred. No. 1.5e-23;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 839 MKNNHSGSTSTLLTLTILHSDGDLTEQKISKPMRSLRLAAGSAIVKLAQEPCHET 898
Db 1 MKNNHSGSTSTLLTLTILHSDGDLTEQKISKPMRSLRLAAGSAIVKLAQEPCHET 60
OY 899 ITLEQYOLCALAINDECYQROYFAOKLHKG 929
Db 61 ITLEQYOLCALAINDECYQROYFAOKLHKG 91
RESULT 14
AAU27999
ID AAU27999 standard; Protein; 111 AA.
AC AAU27999;
DT 18-DEC-2001 (first entry)
XX
DE Human contig polypeptide sequence #152.
XX
XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytoskeletal; antirheumatic; antiallergic; vulnery; antinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteoplastic; antidiabetic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001MO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Dmanac R;
XX
XX
DR WPI: 2001-589862/66.
DR N-PDB; AAS44899.
XX
PS Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
XX Claim 10; Page 144; 153pp; English.
PS
CC Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC contig polypeptides of the invention. The proteins and their associated
CC DNA sequences are useful for the treatment, diagnosis and prevention of
CC various types of disorder in a mammalian subject such as a human, dog,
CC monkey, mouse, hamster or rat. The disorders include cancers such as
CC leukemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernerke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture or to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 111 AA;
OY Query Match 3.8%; Score 276; DB 22; Length 111;
Db Best Local Similarity 100.0%; Pred. No. 9.1e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1337 TEOKSKSKOHVRRAQORAEPPSSAISTGSTPOKGRPKTSPSPKKNV 1391
Db 1 TEOKSKSKOHVRRAQORAEPPSSAISTGSTPOKGRPKTSPSPKKNV 55
RESULT 15
AAB95449
ID AAB95449 standard; Protein; 233 AA.
AC AAB95449;
XX
DT 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:17905.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX

THIS PAGE BLANK (USPTO)

RESULT 1	
AK045159	
LOCUS	AK045159
DEFINITION	Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length- enriched library, clone:B130042B12 product:4910.1.2 (ADROGEN-INDUCED PROSTATE PROLIFERATIVE SHUFFLE ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION	AK045159
VERSION	AK045159.1 GI:26337114
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) MEDLINE 99279253 PUBMED 10349636
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komou, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
TITLE	
JOURNAL	

MEDLINE PUBMED REFERENCE AUTHORS	20499374 11042159 3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, Y., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, K., Ohara, E., Watabiki, M., Onoda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	4 20530913 11076861	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schiraldi, L.M., Staudli, F., Suzuki, R., Tomita, M., Wagner, U., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldirelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyono, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	5 21085660 11217851	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
TITLE JOURNAL REFERENCE AUTHORS	6 (bases 1 to 3313)	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takano-Kahlira, S., Takekida, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hnako Atsugi City,
COMMENT		

FEATURES	source	location/Qualifiers
	1..3313	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="C57BL/6J"
		/db_xref="FANTOM:DB:B130042B12"
		/db_xref="taxon:10090"
		/clone="B130042B12"
		/tissue_type="parthenogenote"
		/clone_1lb="RIKEN full-length enriched mouse cDNA library"
		/dev_stage="9.5 days embryo"
		<1..2476
		/note="unnamed protein product; 49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens] (SPR1096KV4, evidence: FASTV, 92.3%id, 100%length, match=1430) putative"
		/codon_start=2
		/protein_id="BAC32242.1"
		/db_xref="GI:26337115"
		/translation="VMSKIDGTADDEDEGVPTQAIKAGELKLVSTPHISFNSAE TPFSLLAKLMDKEVAEALQIFKNTGSIEDPPIKRIALLPHKRSKQPPRA KVAIHCHAFESSEKTOFQIPEPLKHSIDLSPNLEHLIPVLTGHALIAPAPAP LKSLVATFIYKDLMDRLDPEKTTKLVNVDDEVSPEPTWIKAIKMYVWLGKNN HSGVSTLELTITLHSDDLLEQGTSLKPSKRLKAGSIVKADQPCHEITIT LEOYOLCALAINDECYOVROVFKLHKHSRLRLPEVATICALAKDQKERRRAH RCOLVNIETRYRRLKOHVAASEKLSLPEVYVPTIHLADPDYVKVODIEQLKD VKCLNFEVLEIILAKNNENSHAFIRKVENIKOTKADGPDYKMEKLTVCDDVAM IISKSTYTSLESKDPVLPARFPTQKVENISNTKVLPEPKMSFPPGKPKXTANV ANKPLSSAKQKOTQSSRMETYSNASSSSNPSRIRGRIDSSMDHSENDYAMS SPLPKSKDRKEDPDLVRSLEKPRSKKAPVTDPEKLSMDLTLVDQKKGSR GKRGRGASDSDQDQWPEEKRRHEELLENDQNSPPKGRGRPRPKGLGTSKEEP TMTKTSKGNKKRLVPPVDDDEEERQIGTENTERSKSKQHRTRKRAQRAESPESTAV ESTQSTPQKGRGRSKAPSPSOPPKIRVGRSKVATKENDSSEMDVLAQASPSVD TPOEGAEEDDISGVNRRRSKRERR"
		3288..3323
		/note="putative"
		polyA_signal
		3313
		/note="putative"
		polyA_site
		3313
		/note="putative"
		BASE COUNT 1066 a 709 c 732 g 806 t
		ORIGIN
		Query Match 45.5%; Score 1899.6; DB 11; Length 3313;
		Best Local Similarity 89.2%; Pred. No. 0;
		Matches 2049; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
QY	1872	AGTGAACCAATCATAGATGGAACAGCAGATGATGAAGATGAGGAGTTCACATCATCA 1931
DB	1	AGGTAATAATCATATGATGGAACAGCAGATGATGAAGATGAGGAGTTCACATCATCA 60
QY	1932	AGGCATCAGACAGAGCTTCTGAAGTACTCTCATTTACACATCCATCTCATAT 1991
DB	61	AGCATACAGGCGAGCGCTTGAGCGTGAAGGAGTCTTTACGACACCTATTCCTT 120
QY	1992	TCAATTCGTGAAGACATTTGAATCATTTACTGCTTGCTGAATAATGATGATGAAGA 2051
DB	121	TCACTCGCTGAGACAGCTTGTAGTCCCTTGTGCTGCTGCAATGATGATGATGAAGA 180
QY	2052	AGCAGAGCTGCACCTCAATTTTCAAAAAACAGAGAGCAAAATGAAGAGATTTCG 2111
DB	181	GGCAGAGCTGCCTACAAATTTTCAAAAAACAGAGAGCAAAATGAAGAGATTTCG 240
QY	2112	ACACATCAGATCAGAGCTTGTCTCTGTTTTCATCATCAAAATCTAAAAAGAGACCC 2171
DB	241	TCAATCATGATCAGAGCTTGTCTCTGTTTTCATCATCAAAATCTAAAAAGAGACCC 300
QY	2172	TCAAGCCAAATATGACATTTATTTATTCATGCGATATTTTCTAGTAAAGAGACCA 2231

Db	301	GCAGGCCAAATACCGCATTCATTTGATTCATGCCATATTTTCTAGTAAAGAACCCAGTT	360
Qy	2232	TGCACAGATATTGAGCCTCTGCATAGAGCCTAGATCCAAAGCAACTGGAACATCTCAT	2291
Db	361	TGCACAGATATTGAGCCTCTGCATAGAGCCTAGATCCAAAGCAACTGGAACATCTCAT	420
Qy	2292	AACACATTTGGTACTATTTGGTGCATTTGCTCTCTTGCACCTGATCAATTTGCTGCTCC	2351
Db	421	AACCCCTGCTGACACTATTTGGTGCATTTGCTCTCTTGCACCTGATCAATTTGCTGCTCC	480
Qy	2352	TTGGAAATCTTGGGAGTACTTTCATTTGTAAGAGATCTTGCATGAATGATGGCTTCC	2411
Db	481	TCTGAGCTCTTGGGAGTACTTTCATTTGTAAGAGATCTTGCATGAATGATGGCTTCC	540
Qy	2412	AGGGAAGAAAGACAACTAACTTTGGGTTCCAGATGAGAGAGATGCTCTGAGACAAATG	2471
Db	541	AGGGAAGAAAGACAACTAACTTTGGGTTCCAGATGAGAGAGATGCTCTGAGACAAATG	600
Qy	2472	CAAAATTCAGGCTATTAAATGATGGTGCATGGCTACTTGGAAAGAAATTAATCACAG	2531
Db	601	CAAAATTCAGGCTATTAAATGATGGTGCATGGCTACTTGGAAAGAAATTAATCACAG	660
Qy	2532	TAAATCAGAGACTTCTACCTTAAGATTGCTAAACAACATATTGCATAGTAGAGACTT	2591
Db	661	TAAATCAGAGACTTCTACCTTAAGATTGCTAAACAACATATTGCATAGTAGAGACTT	720
Qy	2592	GACAGAACAGGGGAAATTAAGTAAACAGATATGCTCAGCTGAGACTTGGCTGGAG	2651
Db	721	GACAGAACAGGGGAAATTAAGTAAACAGATATGCTCAGCTGAGACTTGGCTGGAG	780
Qy	2652	TGCAATGTGAGAGCTGGCAGACAAAGACCTGTTACATCAATCAATTAAGACATTA	2711
Db	781	TGCAATGTGAGAGCTGGCAGACAAAGACCTGTTACATCAATCAATTAAGACATTA	840
Qy	2712	TCACATATGTGCTAGTATCAACAGATGAATGCTCAAGTAAAGACAAGTTGGCCCA	2771
Db	841	TCACATATGTGCTAGTATCAACAGATGAATGCTCAAGTAAAGACAAGTTGGCCCA	900
Qy	2772	GAAACTTCACAAAGGCTTTCCCGTTACGGCTTCCACTTGAATATATGCAATCTGTC	2831
Db	901	GAAACTTCACAAAGGCTTTCCCGTTACGGCTTCCACTTGAATATATGCAATCTGTC	960
Qy	2832	CCTTTGTCACAAAGATCCTGTAAAGAGAGAGACCTATGCTAGACAGTGTGGTGA	2891
Db	961	CCTTTGTCACAAAGATCCTGTAAAGAGAGAGACCTATGCTAGACAGTGTGGTGA	1020
Qy	2892	AAATTAATTAATTAAGGCGGAGATCTGAAGCAGCATGACCTGTTAGTGAATAATTT	2951
Db	1021	AAATTAATTAATTAAGGCGGAGATCTGAAGCAGCATGACCTGTTAGTGAATAATTT	1080
Qy	2952	GTCCTTCTACACAGATATGTTTCCATATACATTTACCTTTGGACATGACCCGA	3011
Db	1081	GTCCTTCTTACCAAGATATGTTTCCATATACATTTACCTTTGGACATGACCCGA	1140
Qy	3012	TTATGTCAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGCTTTGGTGT	3071
Db	1141	TTATGTCAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGCTTTGGTGT	1200
Qy	3072	TCTGGAATATTAAATGGCTAAATAATGAATAATACAGTCAAGCTTTTATCAGAAAGAT	3131
Db	1201	TCTGGAATATTAAATGGCTAAATAATGAATAATACAGTCAAGCTTTTATCAGAAAGAT	1260
Qy	3132	AGAAATATTAAACAAACAAAGATGCCAAGGACAGATGATGCAAAATGATGAATA	3191
Db	1261	AGAAATATTAAACAAACAAAGATGCCAAGGACAGATGATGCAAAATGATGAATA	1320
Qy	3192	ACTGTACACTGTGTGATGTTGGCATGATATCATCTGCAAGAGATTAACATACAG	3251
Db	1321	ACTGTACACTGTGTGATGTTGGCATGATATCATCTGCAAGAGATTAACATACAG	1380
Qy	3252	TTTGGAATCTCTAAAGACCGGTACTACAGCTGTCTTCACTCAACCTGACAGAA	3311
Db	1381	CCTGAGCTCTCTTAAGAACCCCGTCTGACAGCTGGTTTTTCACCCAGCCCTGACAGAA	1440
Qy	3312	TTTCAGTAACACCAAAATTTATCTGCTCTGGAATGAATAATCTTTTCACTCTGAAA	3371
Db	1441	TTTCAGTAACACCAAAATTTATCTGCTCTGGAATGAATAATCTTTTCACTCTGAAA	1500
Qy	3372	ACCTAAACCAACCAATGTTCTAGAGAGCTGTTAAACAGGCACCTTTCATCAGCAGAAC	3431
Db	1501	ACCTAAACCAACCAATGTTCTAGAGAGCTGTTAAACAGGCACCTTTCATCAGCAGAAC	1560
Qy	3432	ATCTCAGACCAAAATCATCAGATGAAGAACTTAAGCATGCAAGCAGACAGCTCAATTC	3491
Db	1561	ATCTCAGACCAAAATCATCAGATGAAGAACTTAAGCATGCAAGCAGACAGCTCAATTC	1620
Qy	3492	AAGCTCTCTGGAAGATTAAGGGAGGCTTGATAGTTCTGAAATGATCAGAGTGAATA	3551
Db	1621	AAGCTCTCTGGAAGATTAAGGGAGGCTTGATAGTTCTGAAATGATCAGAGTGAATA	1680
Qy	3552	TGAAGATTACCAATGCTTCACTTGGCCGGGAAAAAAGTGAAGAGAGAGACTGC	3611
Db	1681	TGAAGATTACCAATGCTTCACTTGGCCGGGAAAAAAGTGAAGAGAGAGACTGC	1740
Qy	3612	TGATCTTGAAGGTTGTAATTGAGAACCTTAGAGCAGAGAAAAAACCCGTCACAGA	3671
Db	1741	TGATCTTGAAGGTTGTAATTGAGAACCTTAGAGCAGAGAAAAAACCCGTCACAGA	1800
Qy	3672	ACAGGAGAGAAATTAAGTATGATGATGACTTGACTTAAGTTGGTACAGGAACGAACCTTA	3731
Db	1801	ACAGGAGAGAAATTAAGTATGATGATGATGACTTGACTTAAGTTGGTACAGGAACGAACCTTA	1860
Qy	3732	AGGCGATGAGGAGTGGGAAAGAGGCGCATAGGCTTCAAGATGATGATGACACAGATG	3791
Db	1861	AGGCGATGAGGAGTGGGAAAGAGGCGCATAGGCTTCAAGATGATGATGACACAGATG	1920
Qy	3792	GCTTGAGGAGAAAGAGCTCAAGAGATATATTAGAAATGAAGATGAACGAATATGTC	3851
Db	1921	GCTTGAGGAGAAAGAGCTCAAGAGATATATTAGAAATGAAGATGAACGAATATGTC	1980
Qy	3852	GCCAAAAAAGGTTAAAGAGCCGACACCAAAACCTTTGGTGAAGGTACACCAAAAGA	3911
Db	1981	GCCAAAAAAGGTTAAAGAGCCGACACCAAAACCTTTGGTGAAGGTACACCAAAAGA	2040
Qy	3912	AGAGCCCAATGAAGAACTTGTAAAAAGGAAGCAAAAAAATCTGACCTCCAGCAC	3971
Db	2041	AGAGCCCAATGAAGAACTTGTAAAAAGGAAGCAAAAAAATCTGACCTCCAGCAC	2100
Qy	3972	AGAGGAGAGAGAGAGAGAGAGACAAATGCGAATACGGAACAGAGTCCAAAGCA	4031
Db	2101	AGAGGAGAGAGAGAGAGAGAGACAAATGCGAATACGGAACAGAGTCCAAAGCA	2160
Qy	4032	ACAGCAGCAGTGTCAAGAGAGCAGACAGAGCAAGTCTGTAATCTAGTGCAT	4091
Db	2161	ACAGCAGCAGTGTCAAGAGAGCAGACAGAGCAAGTCTGTAATCTAGTGCAT	2220
Qy	4092	TGAATCCACACAGTCCACACACAGAGAAAGAGAGAGACATCAAAAGCCATCAC	4151
Db	2221	TGAATCCACACAGTCCACACACAGAGAAAGAGAGAGACATCAAAAGCCATCAC	2280
Qy	4152	ATCACACCAAAAAA 4169	
Db	2281	ATCACACCAAAAAA 2298	
RESULT 2			
AK041682			
LOCUS			
DEFINITION			
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched			
library, clone:4630029m15 product:49110.1.2 (ANDROGEN-INDUCED			
PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)			
(FRAGMENT) homolog [Homo sapiens], full insert sequence.			
ACCESSION			
AK041682			
VERSION			
AK041682.1 GI:26334660			
KEYWORDS			
HTC; CAP trapper.			

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042155

REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohtsuka, E., Matsuura, S., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Akawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakawa, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kiehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S., and Hayashizaki, Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851

REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 12108560
PUBMED 12108560

TITLE Muramatsu, M. and Hayashizaki, Y.
JOURNAL Direct Submission
COMMENT Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 2899
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:A630029M15"
/db_xref="taxon:10090"
/clone="A630029M15"
/tissue_type="thymus"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
117. 2222
/note="unlabeled protein product; 49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOPROX 2) (FRAGMENT) homolog [Homo sapiens] (SPR1096KV4, evidence: PASTY, 92.3%ID, 100%length, match=1430) putative"
/codon_start=1
/protein_id="BAC31031.1"
/db_xref="GI:26334661"
/translation="WITRLPPPGKADPFKFTVOLEDEKIRQLALVSPSC KQEGCAREITKRLNPKPTNPFLEIKRLERLAVHIDTESIALRKQVKSIDG TADDEDEGVPTDAIRAGLELVLSTFTHSPHSASETEFLACMKDEKVAEAL QIFRNKSGSKDEEPFIRSLALVPLHRSKSPKQKVAHICLHAFSSKEQFOI FEPHLKSLDPSNLEHTPLVTGILALADPOFAPIKSLVTFIVKDLIMDRDGL KKTTLKLVPPDEVSPTWVTKQIKKMWVRLLMKNNHSGSTIRLITLHSDG LITQGKISKPDMSRLRLAAGSALVLAQEPCTHEITLEIYOLCALAINDECTOYKOV FAQLRHGLSLRLPLEYMAICALADPVKERRARAROCVLRNRYLRKQOHAV SEKLILSLPEYVVPYIHLAHPDYKVDIDLDVKECLMFEVLEIMAKNNNSH AFIRKVENIKOTKDAOGPDCTKNEKLYVCVANNIIMSKSTVLSLESPKQVPA RFPTOPDKPNSNTKNTLPEPMKSEFPNGKRTANVGAVKPLSSAGKOSOTSSRE TVGNASSSSNPSSRGRTKGRLDSEMDHSHNEDYTMSSPLPKKSKRDEPDLVYRC LVGRVRLILVIVLIPAL"

BASE COUNT 868 a 582 c 617 g 832 t
ORIGIN

Query Match 41.9%; Score 1750.2; DB 11; Length 2899;
Best Local Similarity 90.2%; Pred. No. 4.8e-20;
Matches 1872; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 1550 AACGATGCCAGTGTCAAGGCCATATTTCAAAAGTAGTGGTATTACAGAAGATTAC 1609
DB 79 AGACAGATGCCAGTGTCAAGGCCATATTTCTAAAGTAGTGGTATTACAGAAGATTAC 138
QY 1610 CTGATCTGTTAAGCTCAGAGATTTCATGAAGAAATTCACACAGGTTTGAAGATGATG 1669
DB 139 CAGATCTGTTAAGGCCCCAAGACTCATGAAGAAATTCACACAAAGTGTGAAGATGATG 198
QY 1670 AGAAATTAAGAAAGCACTTGAAGTACTGTGTGTCACATGCTCTGCAAGAGCTG 1729
DB 199 AGAAATTAAGCAAGCACTGTAAGCTTGTAGTGTCCACATGCTCTGCAAGAGCTG 258

QY	1730	AAGTTGTGTGCGTAAATAACTAGAAAGTTGGGCAACCCCAACAGCCTACAAATCCTT	1788
Db	259	AAGCGTGTGCGTAAATAACTAGAAAGTTGGGCAACCCCAACAGCCTACAAATCCTT	318
QY	1790	TCCTGGAAATGATCAAGTTTCTCTTGGAGAGATATGCACCTGTGCATATAGATACGAAT	1848
Db	319	TCTTGGAAATGATCAAGTTTCTCTTGGAGAGATATGCACCTGTGCATATAGATACGAAT	378
QY	1850	CTATCAGTGTCTTATTAACAGAGCAATCAATATGATGGAACACACAGATGATGAAG	1908
Db	379	CCATCAGTGTCTTATTAACAGAGCAATCAATATGATGGAACACACAGATGATGAAG	438
QY	1910	ATGAGGGTGTTCACCACTGATCAAGCCATCAGAGCAGCTTGAACTGCTTAAGGACTCT	1968
Db	439	ATGAGGGTGTTCACCACTGATCAAGCCATCAGAGCAGCTTGAACTGCTTAAGGACTCT	498
QY	1970	CATTACACATCCCATCTCATTTCTATTCTGCTGAACACTTGGATCATTTACTGCTTTC	2028
Db	499	CTTTTACCACTCATCTCTCTTCTCATCTGCTGACACCTTTGAGTCCCTTGTGGCTGTTC	558
QY	2030	TGAAATATGATGATGAAAAAGTAGAGAGAGCGCATCAAAATTTTCAAAAACACAGGAA	2088
Db	559	TGAAGATGATGATGAGAGAGTGGCAGAGCTGCATCAAAATTTTCAAAAACACAGGAA	618
QY	2090	GCAAAATTTGAAGAGATTTTCCACACATCAGATCAGCCTTGCCTCTTTTACATCACA	2148
Db	619	GCAAAATTTGAAGAGATTTTCCCTCCTCATCAGATCAGCTTGTCTGTCTACTACATACA	678
QY	2150	AATCTAAAAAGAGACCCCGCTCAGCCCAATATGCCATTTGTATCCATGCGATAT	2208
Db	679	AATCCAAAAAGAGACACCTCGGCAGGCCAAATATGCCATTCATTTGTATCCATGCGATAT	738
QY	2210	TTTTAGTAAAGAGACCCAGTTTGGACAGATATTTGAGCCTCTGCATTAAGACCTAGATC	2268
Db	739	TTTTAGTAAAGAGACCCAGTTTGGACAGATATTTGAGCCTCTGCATTAAGACCTAGATC	798
QY	2270	CAAGCAACCTGGAAACATCTCATTAACCACTTGGTACTATTGGTCATATTCCTCTCTTG	2328
Db	799	CAAGCAACCTGGAGACATCTTATTAACCCCTTGTCATATTGGTCATATTCCTCTCTTG	858
QY	2330	CACCTGATCAATTTGCTGTCTCCTTGGAAATCTTGGGTAGCTACTTTCTTGTGAAGATC	2388
Db	859	CACCTGATCAATTTGCTGTCTCCTCTCTGGAAGCTTTGGTGGCAACTTTCATTTGTGAAGACC	918
QY	2390	TTTCTATTAAGATGCGGCTTCACAGGGAAGAAAGACAACTTAACCTTGGGTTCCAGATGAG	2448
Db	919	TTCTATTAAGATGCGGCTTCACAGGGAAGAAAGACAACTTAACCTTGGGTTCCAGATGAG	978
QY	2450	AAGTATCTCTTGAGCAATGTGTCAAAATTCAGGCTATTAATATGATGTTCGATGCGTAC	2508
Db	979	AAGTCTCACCTGAGAGCAATGTGTCAAAATTCAGGCTATTAATATGATGTTCGATGCGTAC	1038
QY	2510	TTGGAAATGAAAAATATATCAGATAATCAGGAATCTTACCTTAAGATTTGCTTAACACAA	2568
Db	1039	TTGGAAATGAAAAATATATCAGATAATCAGGAATCTTACCCCTGATTTACTAACACAA	1098
QY	2570	TATTCGATGATGATGAGAGACTTGCACAGAAACAGGGGAAAAATATATTAACCAATATGTAC	2628
Db	1099	TATTCGATGATGATGAGAGATTTTGCACAGAAACAGGGGAAAAATATATTAACCAATATGTAC	1158
QY	2630	GTCGTAGACTTGTGCTGGAGTGTATTTGGAAGCTGGACAAGAACCTGTATACATG	2688
Db	1159	GCTGTAGACTTGTGCTGTGGAGTGTATTTGGAAGCTGGACAAGAGCCCTGTATTACACAG	1218
QY	2690	AAATCATCATTTAGAACAAATATCAGCATATGTGATTAAGCTATCAACGATATGCTATTC	2748
Db	1219	AGATCATTTACACTGGAGCAGTACACAGCTGTGTGATTTAGCATCATGATGAGTGTATTC	1278
QY	2750	AAGTAAAGCAAGTGTGTCCCAAGAACTTCAAAAGGCCCTTCCGTTTACGGCTTCCAC	2808
Db	1279	AAGTAAAGCAGGTGTGTCCCTCAGAAACTTCAAAAGGCCCTTCCGTTTACGGCTTCCCC	1338

QY	2810	TTGAGTATATGGCAATCTGTGTCCTTTGTGCAAAAGATCCGTGAAGAGAAAGACTG	2863
Db	1339	TTGAGTACATGGCCATCTGTGCTCTTTGTGCAAAAGACCCGTGAAGAGAGGAGCC	1398
QY	2870	ATGCTAGCAATGTTGGTGGAAAAATTAATTAATGAAGCGGAGAGTATCTGAAGCAGATG	2929
Db	1399	ATGCTAGCAAGTCTTGGTGGAAACATCACTGTGAAGAGGAGTACCTGAAGCAGATG	1458
QY	2930	CAGCTGTTAGTGAATAAATTAATGTCTCTTACACAGAGTATGTTGCCATATCAATTC	2989
Db	1459	CAGCTGTTAGTGAATAAATTAATGTCTCTTACACAGAGTATGTTGCCATATCAATTC	1518
QY	2990	ACCTTTTGGCAACATAGCCACAGATTAATGTCAAAAGTACAGAGATTTGAACACTTAAAGTG	3049
Db	1519	ACCTTTTGGCAACATAGCCACAGATTAATGTCAAAAGTACAGAGATTTGAACACTTAAAGTG	1578
QY	3050	TTAAAGATATGCTTGGTGTGTGTCGGAAATATTAAATGGCTAAATAATGAATAATTAACGTC	3109
Db	1579	TGAAGAATATGCTTGGTGTGTGTCGGAGATATTGATGGCTAAATAATGAATAACAGACC	1638
QY	3110	ACGCTTTATCAGAAAGATGGTAGAATAATTTAAACAACAAGATGCCCAAGACCAG	3169
Db	1639	ATGCACTTATCAGAAATAATGATAGAAATATTAAACAACAAGATGCCCAAGACCAG	1698
QY	3170	ATGATGCAAAATGATGAATAAACTGTACACTGTGTGTGATGTTGCCATGATATCATCA	3229
Db	1699	ATGATGCAAAATGATGAATAAAATTTGTACACCGTGTGTGATGTTGCCATGATATCATCA	1758
QY	3230	TGTCAAAAGATGACTATCATACAGTTGGAAATCTCCAAAGACCCGGTACTTACACACTGGTT	3289
Db	1759	TGTCAAAAGATGACTATCATACAGTTGGAAATCTCCAAAGACCCGGTACTTACACACTGGTT	1818
QY	3290	TCTTCACCTCAACCTGACACAAGATTTTCAGTAAACACCAAAAATTAATCTGCTCTGAAATGA	3349
Db	1819	TTTTCACCCAGCGCTGACACAAGATTTTGTAGTAACACCAAAAATTAACCTGCTCTGAAATGA	1878
QY	3350	AATCATTTTTCACCTCTCTGAAAAACCTTAACACCAATGTTCTAGAGACTTTTAACAGC	3409
Db	1879	AATCATTTTTCACCTCTCTGAAAAACCTTAACACCAATGTTCTGAGACACTTTAATTAAGC	1938
QY	3410	CACTTTCATACAGGAGCAAGCAATCTGACCAATCATCTACGAATGGAATCTGTAAGCA	3469
Db	1939	CACTTTCATACAGGAGCAACAGCTCTGACCAATCATCTACGAATGGAATCTGTAAGCA	1998
QY	3470	ATGCAACGACAGCTCAATCCAAAGCTTCTCTGGAAGATTAAGGAGGAGCTTGATAGTT	3529
Db	1999	ACGCAAGACAGCTCAATCCAAAGCTTCTCTGGAAGATTAAGGAGGAGCTTGATAGCT	2058
QY	3530	CTGAATATGATCACAGTGAATAATGAAGATTACACAAATGCTTTCACCTTTGCCGGGAAAA	3589
Db	2059	CTGAATATGATCACAGTGAATAATGAAGATTATACAAATGCTTTCACCTTTGCCAGGAAAA	2118
QY	3590	AAAGTGACAAAGAGACGACTCTGATCTTGTGAAG	3624
Db	2119	AAAGTGACAAAGAGAGAACCCCTGATCTTGTGAAG	2153
RESULT 3			
AK086753			
LOCUS	AK086753	4273 bp	mRNA
DEFINITION	Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:DJ30049H23 Product:45J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOPRM 2)		
ACCESSION	AK086753		
VERSION	AK086753.1	GI:26103739	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	Carninci, P. and Hayashizaki, Y.		

JOURNAL MEDLINE PUBMED REFERENCE	High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636
AUTHORS	2. Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
JOURNAL MEDLINE PUBMED REFERENCE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20495374 11042159
AUTHORS	3. Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, T., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, R., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T., Kasliwal, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL MEDLINE PUBMED REFERENCE	sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
AUTHORS	4. Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Salto, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Salto, R., Kadota, K., Matsuda, H., Ashburner, M., Batelov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Koehwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudil, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kanaya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsuki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED REFERENCE	Nature 409 (6821), 685-690 (2001) 21085660 11217851
AUTHORS	5. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL MEDLINE PUBMED REFERENCE	of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 4273)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Hara, A., Hashizume, W., Fukuda, S., Furuno, M., Hanagata, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hoti, R., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsumi, N., Okazaki, Y., Salto, R., Salto, H., Sakai, G., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama

COMMENT	FEATURES	source
kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp), URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)		
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details.		
URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/. Location/Qualifiers		
1. .4273		
/organism="Mus musculus"		
/mol_type="mRNA"		
/strain="C57BL/6J"		
/db_xref="FANTOM,DB:D930049H23"		
/db_xref="taxon:10090"		
/clone="D930049H23"		
/tissue_type="head"		
/clone_id="RIKEN full-length enriched mouse cDNA library"		
/dev_stage="15 days embryo"		
<1.2245		
/note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens] (SPTR1Q96KV4, evidence: FASTV, 92.3%ID, 100%length, match=1430) putative"		
BASE COUNT 1363 a 840 c 899 g 1171 t		
ORIGIN		
Query Match 39.7%, Score 1655.2, DB 11, Length 4273; Best Local Similarity 88.8%, Pred. No. 9.2e-274; Matches 1836, Conservative 0; Mismatches 228; Indels 4; Gaps 4;		
OY 2104 GATTTTCACACATCAGATGACGCTTCGTCCTCTTTTACATCAACAATCT-AAAAAGG 2162		
DB 1 GATTTCCTCTACATCAATCAAGCTTCTTCTACTACATCAACAATCAAAAAAG 60		
OY 2163 ACCCCCCCGTCAGCCAAATATGCGATTCATTGTATCGATGCGATATTTCTAGTAAGA 2222		
DB 61 ACCACCTTCGCGAGCGCCAAATACGCCATTCATTGTATTCATGCCATATTTCTAGTAAGA 120		
OY 2223 GACCAGTTTSCACAGATATTTGAGCGCTGSCATTAAGAGCCTAGATCCAGCAACCGGA 2282		
DB 121 AACCCAGTTTSCACAGATATTTGAGCCCTGSCATTAAGAGCCTAGATCCAGCAACCGGA 180		
OY 2283 ACATTCATATAACACATTTGTTACTATTTGGTCATATTTGCTCTCCCTTCACCCGTATCAAT 2342		
DB 181 GCATCTTATAACCCCTCTGTCGACTATTTGGTCATATTTGCTCTCTCCACCTGATCAGTT 240		
OY 2343 TGCATGCTCTGGAAATCTGGATAGCTACTCTCATTTGTGAAGATCTTCATGAATGA 2402		
DB 241 CGCTGCTCTCTGGAAGTCTTTGGGCGCACTTTCATTTGTGAAGAGCCCTCCATGAATGA 300		
OY 2403 TCGGCTTCACAGGAAAAAGACAACTTAACCTTTGGGTTCCAGATGAAGAATGATCTCCGTA 2462		
DB 301 CCGGCTTCACAGGAAAAAGACAACTTAAGCTTTGGGTTCCAGATGAGGAAGTCACTCCTGA 360		
OY 2463 GACATATGTCAAAATTCAGGCTATTTAAATGATGTTGATGCTTACTTGGAAATGAAAA 2522		
DB 361 GACATATGTCAAAATTCAGGCTATTTAAATGATGTTGATGCTTACTTGGAAATGAAAA 420		
OY 2523 TAATTCACAGTAAATCAGGAATCTTACTCTTAAGTTGTATCAACAACTATTTGCATAGGA 2582		
DB 421 TAATTCACAGTAAATCAGGAATCTTACTCTTAAGTTGTATCAACAACTATTCATAGGA 480		
OY 2583 TGGAGACTTGCACAGAACAGGGAAATTTAGTAAACCGAGATATGTCAGTCTGAGACTTGC 2642		
DB 481 TGGGGAATTTGCACAGAACAGGGAAATTTAGTAAACCGAGATATGTCAGGCTGAGACTTGC 540		
OY 2643 TGTCTGGAGTCTCTTTGTGAAGCTTGGCAACAAGAACCTGTTACATGAATCATACATT 2702		

Db	541	TGCTGGAGTGTCTATTGTGAAAGCTGGCAGAGAGCCGTTTACACGAGATCATTTACCT	600
OY	2703	AGAAACAATATAGCTATGTGCTTATAGTATCAACGATCAATGCTTATCAAGTAAGCAAGT	2762
Db	601	GGAGAGATACAGCGTGTGTGATTAGGCATCAATGATATAGTGTATTAAGTCAGGACAGT	660
OY	2763	GTTTGCCCAAACTTCACAAAAGGCGTTTCCGTTTACGGCTTCCACTT - GAGTATATGG	2821
Db	661	GTTGCGTCAGAAACTTCAAAAGGCGTTTCCGCTTACGGCTTCCCTTGGAGTACATGG	720
OY	2822	CAATCTGGCCCTTGTGTCAAAATCTCTGTAAAGGAGAGAAAGAGCTATCTAGGCAAT	2881
Db	721	CCATCTGTGCTTGTGTGCCAAAGCCCTGTGAAAGAGAGGGAGCCCATCTGTAGACGT	780
OY	2882	GTTTGGTGAATAATATTAATATGTAAGGGCGGAGTATCTAAGCAGATCAGCTGTATGTG	2941
Db	781	GCTTGCTGAAGACAT - CCGGTGAGAGGGAGTACCTTAAGCAGATCAGCTGTATGTG	839
OY	2942	AAAAATTTTGTCTCTTCTTACACAGATATGTTTCCATATACAAATTCACCTTTGGCAC	3001
Db	840	AAAAATTTTGTCTCTCTTCTTACACAGATATGTTTCCATATACAAATTCACCTTTGGCAC	899
OY	3002	ATGACCCAGATATATGTCAAACTACAGATATTTGAACAACCTTAAGATGTTTAAGAAATTC	3061
Db	900	ATGACCCAGATATATGTCAAACTACAGATATTTGAACAACCTTAAGATGTTGAAGATATCC	959
OY	3062	TTTGGTTTGTCTGTGAATATTTAATAGGTAAATAATGAATAATACAGTACGCTTTTATCA	3121
Db	960	TTTGGTTTGTCTGTGAGATATTTGATGGCTTAATAATGAATAACAGCCATGATTTATCA	1019
OY	3122	GAAGATGTTAGAAAAATTTAACAACAACAAGATGCCAAGACAGATATGCAAAAA	3181
Db	1020	GAATAATGTTAGAAAAATTTAACAACAACAAGATGCTCAAGACAGATATGTTCAAAAA	1079
OY	3182	TGAATGAAAAACTGTACACTGTGTGTATGTTGCCATATATCATCATGTCAAAAGAGTA	3241
Db	1080	TGAATGAAAAATTTGTACACCGTGTGTGTGTTGCCATATCATCATGTCTCAAGAGCA	1139
OY	3242	CTACATACAGTTTGGAAATCTCTCTTAAGACCCGGTACTACAGCTGCTTCTTCACTCAC	3301
Db	1140	CCACGTACAGCTTGAGTCTCTCTTAAGACCCCGTCTGACAGCTGGTTTTCACCCAGC	1199
OY	3302	CTGACAAGAATTTAGTAAACACAAAAATTTATCTGCTCTCTGGAATGAATCATTTTCA	3361
Db	1200	CTGACAAGAATTTAGTAAACACAAAAATTTACGTGCTCTCTGGAATGAATCATTTTCA	1259
OY	3362	CTCCTGGAAAACTTAACCAACCAATGTTTCTAGAGCTGTTTAACAGCCACTTTCATCAG	3421
Db	1260	CTCCTGGAAAACTTAACCAACCAATGTTTCTGAGAGCTTAAATTAAGCCACTTTCATCAG	1319
OY	3422	CAGGCAACCAATCTCAGACCAAAATCATACGAATGGAACGTGAAGCAATCAAGCAGCA	3481
Db	1320	CAGGCAACCAATCTCAGACCAAAATCATTAAGAAATGGAACGTGAGCAACCAAGCAGCA	1379
OY	3482	GCTCAAAATCAAGGCTCTCTCTGGAAGAAATTAAGGGGAGGCTGATAGTTCTGAATGATCT	3541
Db	1380	GCTCAAAATCAAGGCTCTCTCTGGAAGAAATTAAGGGGAGGCTGATAGTTCTGAATGATCT	1438
OY	3542	ACAAGTGAATAATGAAGATTACACAATGTCTTACCTTTCCGGGAAAAAAAGTGAACA	3601
Db	1439	ACAAGTGAATAATGAAGATTATTAACAATGTCTTACCTTTCCAGGAAAAAAAGTGAACA	1498
OY	3602	GAGAGCACTGTGAATCTTGTAAAGTGTGAATTTGGAAGGCCATGAGGCAAGAAAAAGCC	3661
Db	1499	GAGAGCACTGTGAATCTTGTAAAGTGTGAATTTGGAAGGCCATGAGGCAAGAAAAAGCC	1558
OY	3662	CCGTACACGAACAGAGAGAAAAATTAAGTATGATGATCTTACTAATGTTGGTACAGAAC	3721
Db	1559	CTGTACACAGACCTTGAAGAGAAATTAAGTATGATGATGATCTTACTAATGTTGGTACAGAAC	1618
OY	3722	AGAAACCTTAAGGACAGTACGGAAGTGGAAAAAGGCCATACGGCTTACAGATCTGATG	3781
Db	1619	AGAAACCTTAAGGACAGTACGGAAGTGGAAAAAGGCCATACGGCTTACAGATCTGATG	1678

OY		3782	AACAGCAGTGGCCCTGAGGAAAGAAGCGCTCAAAAGATATATTGTGAATAATGACATCAAC	3841
Dd		1679	AGCAGCAGTGGCCCTGAGGAAAGAAGCGCCAAAGAGGAGACTTCCTGGAAAATGAGCATGAGC	1738
OY		3842	AGAAATAGTCGCCCAA AAAAAGGGTTAAAAAGAGGCGCACCAACCAAAACCTTGTTGGTAGGTA	3901
Dd		1739	AGAACAGCCCAACCAAAAAGGGCAAAAAGAGGACAGCCACCACAACCTTTGTGGGGGA	1798
OY		3902	CACCAAAAAGAGAGCCCAAACTGAATAAATCTTTAAAAAGAAAGCAAAAAAAAAAATCTGCAC	3961
Dd		1799	CATCGAAGAGAAAGAGCCCAATGAAATCATTCAGAAAGAAAGAAACAAAGAAAAAATCTGTAC	1858
OY		3962	CTCCAGCACCAGAGAGGAGGAGAAGAAGAACAAATGTGAATTCGAGACAGAGT	4021
Dd		1859	CTCTGTGTGTGTGATGAGCAGCATGAMAACAAAGAAACAAATTTGGAAACACAGAACATAGT	1918
OY		4022	CCAAAAGCAAAACAGCAGCAGTGTCCAAGGAGAGACACAGCAGAGCAAGAAATCTCTGAA	4081
Dd		1919	CAAAAAGCAAAACAGCAGCAGTGTCCAAGGAGAGACACAGCAGAGCAAGAAATCTCTGAA	1978
OY		4082	CTAGTGCATATGAAATCCACACAGTCCACACGCAAGAAAGAGAGCAAGACATCAAAAA	4141
Dd		1979	CAAGTGCAGTTGAAATCCACACAGTCCACACGCAAGAAAGAGAGCAAGACATCAAAAA	2038
OY		4142	GCCCATCATCCATCCACACCAAAAAAA 4169	
Dd		2039	CACCATCATCCATCCACCAAAAAAA 2066	
RESULT 4				
LOCUS	AKOJ2384			
DEFINITION	AKOJ2384	4427 bp	mRNA	linear
	Mus musculus adult male olfactory brain cDNA, RIKEN full-length			HTC 05-DEC-2002
	enriched library, clone:6430530P15 product:49j10.1.2			
	(ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED			
	PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert			
	sequence.			
VERSION	AKOJ2384			
KEYWORDS	AKOJ2384.1 GI:26082814			
SOURCE	HTC; CAP trapper.			
ORGANISM	Mus musculus (house mouse)			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE				
AUTHORS	1 Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE				
AUTHORS	2			
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
JOURNAL	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to			
PUBMED	prepare full-length cDNA libraries for rapid discovery of new genes			
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)			
AUTHORS	3			
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,			
JOURNAL	Kono, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,			
MEDLINE	Sunji, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,			
PUBMED	Yamanoto, K., Matsumoto, H., Sakauechi, S., Ikegami, T., Kashiyagi, K.,			
REFERENCE	Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,			
AUTHORS	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, T.,			
TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format			
MEDLINE	sequencing pipeline with 384 multicapillary sequencer			
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)			
REFERENCE	4			
AUTHORS	JOURNAL MEDLINE PUBMED			
TITLE	20530913 11076661			

AUTHORS

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, M., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakata, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Guscini, S., Hill, D., Hofmann, C., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL
MEDLINE
PUBMED
21085660
11217851

REFERENCE

AUTHORS

TITLE

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS

6 (bases 1 to 4427)
Adachi, T., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numata, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

Location/Qualifiers

1..4427

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM.DB:6430530P15"

/db_xref="taxon:10090"

/clone="6430530P15"

/sex="male"

/tissue_type="olfactory brain"

/clone_id="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

<1..2062

/note="4910.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog

[Homo sapiens] (SPR1096KV4, evidence: FASTV, 92.3%ID, 100%length, match=1430)
putative"
polyA_signal 4406..4411
polyA_site 4427
/note="putative"
BASE COUNT 1423 a 832 c 919 g 1253 t
ORIGIN

Query Match 36.4% Score 1521; DB 11; Length 4427;
Best Local Similarity 88.3% Pred. No. 9,4e-251;
Matches 1674; Conservative 0; Mismatches 210; Indels 7; Gaps 2;

2279 TGGAAATCATCAACACCATGGTACTGATATGCTATGCTGCTGACACGATC 2338
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 TGGAGATCATTAATACCCCTGCTGACATATGCTATGCTGCTGCGACCGATC 60
Db
2339 AATTGCTGCTCTTGGGAAATCTTGGGAGCTACTTTCATTTGGAAGATCTTTCATGA 2398
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 AGTTCCGCTGCTCTGAAAGTCTTGGTGCACATTTTCATTTGGAAGACCTTCATGA 120
Db
2399 ATGATGGCTTCAGGGAAGAAAGACAATTAATTTGGGTTCCAGATGAAGATATC 2458
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 ATGACCGGCTTCAGGGAAGAAAGACAATTAATTTGGGTTCCAGATGAAGATC 180
Db
2459 CTGAGCAATGTCATAAATTCAGGCTATTAATTAATGATGTTCCAGATGCTTGAATGA 2518
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 CTGAGCAATGTCATAAATTCAGGCTATTAATTAATGATGTTCCAGATGCTTGAATGA 240
Db
2519 AAAATATCAGCTAATTAATCAGGATCTTACCTTAAGATTCATTAACAAATTTGCATA 2578
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 AGATATATCAGCTAATTAATCAGGATCTTACCTTAAGATTCATTAACAAATTTGCATA 300
Db
2579 GTGATGAGCTTCAGGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2638
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GTATGAGGCTTCAGGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 360
Db
2639 TTGCTGCTGAGGATGCTATTTGTAAGCTGACACAGAACCTTTCATTAATTAATTAATTA 2698
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 TTGCTGCTGAGGATGCTATTTGTAAGCTGACACAGAACCTTTCATTAATTAATTAATTA 420
Db
2699 CATTAGAACATATCAGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 2758
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 CACTGGAGCTATCAGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 480
Db
2759 AAGTGTGGCCAGAACTCAGAAAGGCTTCCGTTACGCTTCCAGCTTCCAGCTTCCAGCTT 2818
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 AGGTGTGGCCAGAACTCAGAAAGGCTTCCGTTACGCTTCCAGCTTCCAGCTTCCAGCTT 540
Db
2819 TGGCAATCTGCTGCTTGTGCAAAAGATCTGTAAGAGAGAGAGATCATGCTAGGC 2878
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 TGGCAATCTGCTGCTTGTGCAAAAGATCTGTAAGAGAGAGAGATCATGCTAGGC 600
Db
2879 AATGTTGGTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2938
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 AGTGTGGTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
Db
2939 GTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2998
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 GTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
Db
2999 CACATGACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3058
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721 CACATGACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
Db
3059 GTCTTGGTGTGCTGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3118
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 GCTTGGTGTGCTGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
Db
3119 TCAGAAAGATGTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3178
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
841 TCAGAAAGATGTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
Db


```

OY 3173 ATGCAAAATGATGAAAACTGTACAGTGTGTGATGTGGCATGAATATCATCATGT 3322
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ATACAAAATGATGAAAAATTTGTACACCGTGTGTGATGTGGCATGAATCATCATGT 420
OY 3233 CAAGAGTACTATCATAGTTTGGATCTCTCTTAAACCCCGTACTACACAGCTCTTCT 3292
    || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 CCAGAGGACACAGTACACCTGTCTCAAGAGACCCCGTGTGCGACGCTGGTTT 480
OY 3293 TCACCTACACCTGACAGAAATTTTCAGTACACCAAAATTTATCTCTCTGAAATTAAT 3352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TCACCCAGCTGACAGAAATTTAGTACACCAAAATTTATCTCTCTGAAATTAAT 540
OY 3353 CATTTTCACTCTGAAAAAACCCTAAACCAACATTTCTAGAGCTGTTTACAGCCAC 3412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 CATTTTCACTCTGAAAAAACCCTAAACCAACATTTCTGAGCAGTTAATTAACCCAC 600
OY 3413 TTTTCATCAGGACGACCAATCTCAGACCAAAATCATCAGAAATGAAATCTGAACATG 3472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 TTTTCATCAGGACGACCAATCTCAGACCAAAATCATCAGAAATGAAATCTGAACATG 660
OY 3473 CAAGCAGACGCTCAAAATCCAGCTCTCTGGAAGAAATTAAGGAGGCTTGATAGTTCTG 3532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CAAGCAGACGCTCAAAATCCAGACCTCTCTGGAAGAAATCAAGGAGGCTTGATAGCTCTG 720
OY 3533 AAATGATCAGACAGTGAATTAAGATTTACCAATGTCTTCACTTTGCCGGGAAAAAAA 3592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 AAATGATCAGACAGTGAATTAAGATTTATTCATATGTCTTCACTTTGCCGAAAAAAA 780
OY 3593 GTGACAGAGAGACGACCTGTGATCTTGTAAAGTCTGTAATTTGGAAAGCCTAGAGCAGA 3652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 GTGACAGAGAGAGACGACCTGTGATCTTGTAAAGTCTTCACTTTGCCGAAAAAGCTGGA 834
OY 3653 AAAAAACGCGCCGACAGAAACAGAGAGAAATTTAGTATGATGATGATGATGATGATGATG 3712
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 835 AAAAAACGCGCCGACAGAAACAGAGAGAAATTTAGTATGATGATGATGATGATGATGATG 894
OY 3713 TACAGAGACAGAAATCTAAAGGACGTACGGAAGTCTGGAAGAGGACCATACGCTTCAG 3772
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 895 TACAGAGACAGAAATCTAAAGGACGTACGGAAGTCTGGAAGAGGCGGACCATACGCTTCAG 954
OY 3773 AAATGATGACAGACGAGTGGCTGAGAAAAAGAGGCTCAAAAGAAATATTTAGAAAAATG 3832
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 955 ACTCAGACGAGACGAGTGGCTGAGAGAGAGAGCAAAAGAGAGCTCTGGAATG 1014
OY 3833 AAGATGACAGAAATAGTCCGCAAAAAAGGAGTAAAGAGGCGCAGCAACCAAAACCTTTG 3892
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1015 AAGATGACAGAGAAACGCAACCAAAAAAGGAGCAAAAGAGGCGCAGCAACCAAAACCTTTG 1074
OY 3893 GTGAGAGTACCAAAAGAGGACCAATGAAACTTCTTAAAAAGAGAGCAAAAAA 3952
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 GTGAGAGTACCAAAAGAGGACCAATGAAACTTCTTAAAAAGAGAGCAAAAAA 1134
OY 3953 AAATGATGACAGACGAGTGGCTGAGAAAAAGAGGAGAAAGAAACCAAAAGTGAATTAACG 4012
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1135 AAATGATGACAGACGAGTGGCTGAGAGAGAGAGCAAAAGAGAGCTCTGGAATG 1194
OY 4013 AACAGAGTCCAAAAAGCAAAACAGACGAGTGTCAAGAGAGAGACAGAGAGAGAGAT 4072
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1195 AACATAGTCAAAAGAGCAAAACAGACGCGCATCAAAAGAGAGCAAAAGAGAGAGAT 1254
OY 4073 CTCCTGAATCTAGTCAATTTGAATTCACACAGTTCACACCAAGAAAGAGAGAGAGAC 4132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1255 CTCCTGAATCTAGTCAATTTGAATTCACACAGTTCACACCAAGAAAGAGAGAGAGAC 1314
OY 4133 CATCAAAAAGCGCATCATCATCAACCAAAAAA 4169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1315 CATCAAAAAGCGCATCATCATCAACCAAAAAA 1351

```

```

RESULT 6
BF509252 791 bp mRNA linear EST 06-DEC-2000
LOCUS BF509252
DEFINITION UI-H-B1d-aow-c-07-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone

```

```

IMAGE:3086220 3', mRNA sequence.
ACCESSION BF509252
VERSION BF509252.1 GI:11592550
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 791)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapsr@emil.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a donated poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/ILMIL at:
www.bio.lnln.gov/bdip/image/image.html
Seq primer: M13 Forward
POLYA=yes.
FEATURES
location/Qualifiers
1..791
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3086220"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_Sub8"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I. Site 2: Eco RI. NCI_CGAP_Sub8
is a subcloned library derived from NCI_CGAP_Sub5. The
NCI_CGAP_Sub8 library had 2.5 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub5 was used
as a tracer in a subtractive hybridization with a driver
comprising a pool of clones from NCI_CGAP_Sub5 (IMAGE
clone ids 2732833-2737415, 3068040-3069191; 25% of the
driver population). A pool of clones from NCI_CGAP_Sub4
(IMAGE clone ids 2723592-2729326; 25% of the driver
population), NCI_CGAP_Sub6 (pool Air-AU, IMAGE ids
2728969-2733190; 25% of the driver population), and
NCI_CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550
; 25% of the driver population). Subtraction was
performed as previously described (Bonaldo, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches to Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB=NCI_CGAP_Lu5
TAG_TISSUE=Lung
TAG_SRO=CAAC"
BASE COUNT 270 a 151 c 144 g 226 t
ORIGIN
Query Match 17.2%; Score 716; DB 10; Length 791;
Best Local Similarity 99.6%; Pred. No. 1e-112;
Matches 749; Conservative 0; Mismatches 0; Indels 3; Gaps 3;
OY 2680 TGTACCAGTAATCATCATTAAGAAACAATATACGCTATGTGCAATAGTATCAAGAT 2739
Db 13 TGTACCAGTAATCATCATTAAGAAACAATATACGCTATGTGCAATAGTATCAAGAT 72
OY 2740 GAATGCTATCAATAGACAAAGTGTGGCCCAAAACTTCACAAAGGCTTCCGCTTA 2799
Db 73 GAATGCTATCAATAGACAAAGTGTGGCCCAAAACTTCACAAAGGCTTCCGCTTA 132
OY CGGCTTCACTAGTATATGCAATCTGTGCTTGTGCAAAAGATCTGTAAGAGAG 2859
Db 133 CGGCTTCACTAGTATATGCAATCTGTGCTTGTGCAAAAGATCTGTAAGAGAG 192
OY 2860 AGAAGAGCTCATAGGCAATGTTGTGTAATAATTAATTAAGAGGAGATATCTG 2919

```

```

|||||
Db 193 AGAAGAGCTCATGTAGGCAATGTTGGTGAATAATATTAAGCGGGGATATCTG 252
OY 2920 AAGCAGCATGCAGCTGTTAGTGAATAATATTGCTCTTACACAGATATGTTGCCA 2979
Db 253 AAGCAGCATGCAGCTGTTAGTGAATAATATTGCTCTTACACAGATATGTTGCCA 312
OY 2980 TATTAATTTACCTTTGGCAGATGACCCAGATTATGCAAGTACAGATATGACAA 3039
Db 313 TATTAATTTACCTTTGGCAGATGACCCAGATTATGCAAGTACAGATATGACAA 372
OY 3040 CTTAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATATTAATGCTAAATGAA 3099
Db 373 CTTAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATATTAATGCTAAATGAA 432
OY 3100 AATAACATGCAGCTTTTATTCAGAAAAGATGTAGAAAATATTAACAAACAAAGATGCC 3159
Db 433 AATAACATGCAGCTTTTATTCAGAAAAGATGTAGAAAATATTAACAAACAAAGATGCC 492
OY 3160 CAAGACCATGATGATGCAAAAATGATGAAAACATGTACACTGTGTGATGTGGCATG 3219
Db 493 CAAGACCATGATGATGCAAAAATGATGAAAACATGTACACTGTGTGATGTGGCATG 552
OY 3220 AATATCATCATGCTCAAAAGATGACTACATACAGTTTGAATCTCTAAAGACCCGGTACTA 3279
Db 553 AATATCATCATGCTCAAAAGATGACTACATACAGTTTGAATCTCTAAAGACCCGGTACTA 612
OY 3280 CCAGTGTCTTCTTACTCAACCTGACAAAGATTTTCTAGTAACACCAAAATATATGCTCT 3339
Db 613 CCAGTGTCTTCTTACTCAACCTGACAAAG-ATTTCAGTACACCAAAATATATGCTCT 671
OY 3340 CCTGAATGAATCATCTTTTCTACTCTCTGGAAGAACTTAAACCAACCAATGTTAGAGCT 3399
Db 672 CCTGAATGAATCATCTTTTCTACTCTCTGGAAGAACTTAAACCAACCAATGTTAGAGCT 730
OY 3400 GTTAACAGCCACTTTCATCAGCAGGCAAGCA 3431
Db 731 GTTAACAGG-CACCTTCATCAGCAGGCAAGCA 761

RESULT 7
CA980171 926 bp mRNA linear EST 06-JAN-2003
LOCUS AGENCOURT_11295822 NIH_MGC_164 Mus musculus cDNA clone
DEFINITION IMAGE:30145391 5', mRNA sequence.
ACCESSION CA980171
VERSION CA980171.1 GI:27512825
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 926)
AUTHORS NIH-MGC http://mgi.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabrs-remail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDAM0059 row: f column: 24
High quality sequence stop: 620.
Location/Qualifiers
1..926
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30145391"

```

```

/lab_host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_164"
/note="vector: PCMV-SPORT6.1.ccd, Site_1: EcoRV, Site_2:
NotI; Non-normalized full-length enriched library from
pooled mouse embryonic limb, maxilla and mandible, day
10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp. Average insert size 1.8k bp. Priming
sequence: 5'-GACTAGTCTTCTAGTCCGAGCGGCCGCTT-3'. Tissue
contributed by, David Rowe. Library constructed by Resgen,
Invitrogen Corp."

BASE COUNT      332 a      199 c      222 g      172 t      1 others
ORIGIN

Query Match      16 6%; Score 691.6; DB 14; Length 926;
Best Local Similarity 87.6%; Pred. No. 1.5e-108;
Matches 790; Conservative 0; Mismatches 109; Indels 3; Gaps 3;

OY 3042 TAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATATTAATGCTAAATGAAA 3101
Db 25 TAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATATTAATGCTAAATGAAA 84
OY 3102 TAACAGTACGCTTTTATTCAGAAAAGATGTAGAAAATATTAACAAACAAAGATGCCA 3161
Db 85 CAACAGCATGATTTATTCAGAAAATGTTAGAAAATATTAACAGACAAAGATGCTCA 144
OY 3162 AGGACCATGATGTAAGAAAATGTAAGAAAATGTAAGTACCTGTGTGATGTTGCCATGA 3221
Db 145 AGGACCATGATGTAAGAAAATGTAAGAAAATGTAAGTACCTGTGTGATGTTGCCATGA 204
OY 3222 TATCATCATGCTCAAAAGATGACTACATACAGTTTGAATCTCTTAAAGACCCGGTACTA 3281
Db 205 CATCATCATGCTCAAAAGATGACTACATACAGTTTGAATCTCTTAAAGACCCGGTACTA 264
OY 3282 AGCTGTTTCTTACTCAACCTGACAGAAATTTCACTAACACCAAAATATATGCTCTCC 3341
Db 265 AGCTGTTTCTTACTCAACCTGACAGAAATTTCACTAACACCAAAATATATGCTCTCC 324
OY 3342 TGAATGTAATCATTTTCTACTCTGGAAGAACTTAAACCAACCAATGTTCTGAGCTGT 3401
Db 325 AGAATGTAATCATTTTCTACTCTGGAAGAACTTAAACCAACCAATGTTCTGAGCTGT 384
OY 3402 TAACAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAATCATCAGCAATGAAA 3461
Db 385 TAATAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAATCATCAGCAATGAAA 444
OY 3462 TGTAAAGCAATGCAAGCAGCTCAAAATTCAGTCTCTCTGGAAGAAATTAAGGAGCT 3521
Db 445 TGTGAGCAACGCAAGCAGCTCCACCAAGCTCTCTGGAAGAAATTAAGGAGCT 504
OY 3522 TGATAGTCTGTAAGTGAATGCAAGTCAAGTGAATTAACCAATGCTTCACTTGTGCC 3581
Db 505 TGATAGTCTGTAAGTGAATGCAAGTGAATTAACCAATGCTTCACTTGTGCC 563
OY 3582 GGGGAAAAAAGTGAAGAGAGAGAGAGCTGATCTTGTGAAGTGTGAATGGAAGAGC 3641
Db 564 AGGAAAAAAGTGAAGAGAGAGAGAGAGAGCTGATCTTGTGAAGTGTGAAGTGAAGAGC 623
OY 3642 TAGAGCAGGAAAAAAGCCCGCTCAGACAGAACAGAGGAAAAATTAAGTATGATGCTT 3701
Db 624 TAGAGCAGGAAAAAAGCCCGCTCAGACAGAACAGAGGAAAAATTAAGTATGATGCTT 683
OY 3702 GACTAGTGTGTAAGGAAACAGAACTTAAAGGAGCTGACGCAAGTCCG-AAAAAGGCC 3760
Db 684 AACTAGTGTGTAAGGAAACAGAACTTAAAGGAGCTGACGCAAGTCCG-AAAAAGGCC 743
OY 3761 ATACGCTTCAAGTCTGATGAACAGAGTGTGCTGAGGAAAAAGGCTCAAGAAAGTA 3820
Db 744 GTACAGCTTCAAGTCTGATGAACAGAGTGTGCTGAGGAAAAAGGCTCAAGAAAGTA 803
OY 3821 TATTAGAAATGAAGATGAACGAATTAAGTCCGCAAAAAAGGTTAAAGAGCCGACAC 3880
Db 804 TCCTGAAATGAAGATGAACGAATTAAGTCCGCAAAAAAGGTTAAAGAGCCGACAC 863

```


FEATURES
source

Location/Qualifiers
1..692
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S22SN16n1-50-A08"
/sex="F"
/tissue_type="Ascltes"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/clone_id="S22SN16n1"
/note="Organ: Stomach; Vector: pTY73-Pac; Site:1: EcorI;
Site:2: NotI; The S22SN16 library was contributed by the
Soares laboratory and it was constructed as described by
Bonaldi, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

BASE COUNT 264 a 136 c 148 g 144 t
ORIGIN

Query Match 16.2%; Score 677; DB 12; Length 692;
Best Local Similarity 99.3%; Pred. No. 4.9e-106;
Matches 680; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3186 TGAATACTGTACACTGTGTGTGTGTCATGTAATTCATCATGTCAGAGTACTAC 3245
DB 1 TGAATACTGTACACTGTGTGTGTGTCATGTAATTCATCATGTCAGAGTACTAC 60
OY 3246 ATACAGTTTGAATTCCTTAAGACCGGTACTACACTGCTTCTCCTCCTCAGCCTCGA 3305
DB 61 ATACAGTTTGAATTCCTTAAGACCGGTACTACACTGCTTCTCCTCCTCAGCCTCGA 120
OY 3306 CAGAGATTTAGTAACACCAAAATATCTGCTCTCTGAAATGAATCATTTTCTCCTCC 3365
DB 121 CAGAGATTTAGTAACACCAAAATATCTGCTCTCTGAAATGAATCATTTTCTCCTCC 180
OY 3366 TGGAAACCTTAACACCAATGTTTGAAGAGCTGTAAACAAGCCATTTCTCAGCAGG 3425
DB 181 TGGAAACCTTAACACCAATGTTTGAAGAGCTGTAAACAAGCCATTTCTCAGCAGG 240
OY 3426 CAAGCAATCTCAGACCAATATCATCAGCAATGGAACCTGTAAGCATGCAAGAGACGCTC 3485
DB 241 CAAGCAATCTCAGACCAATATCATCAGCAATGGAACCTGTAAGCATGCAAGAGACGCTC 300
OY 3486 AATTCAGAGCTCTCTGAGNAATTAAGGGAGGCTGTAGTATGTTCTGAATGATCAGAG 3545
DB 301 AATTCAGAGCTCTCTGAGNAATTAAGGGAGGCTGTAGTATGTTCTGAATGATCAGAG 360
OY 3546 TGAATAAGAGATTACCAATGTCTTCACTTTGCCGGGAAAAAAAGTGCACAGAGAGA 3605
DB 361 TGAATAAGAGATTACCAATGTCTTCACTTTGCCGGGAAAAAAAGTGCACAGAGAGA 420
OY 3606 CGACTCTGATCTTGAAGAGTGTGAATTTGAGAGAGCTAGAGGCAAGAAAAAGCCCGCT 3665
DB 421 CGACTCTGATCTTGAAGAGTGTGAATTTGAGAGAGCTAGAGGCAAGAAAAAGCCCGCT 480
OY 3666 CACAGAACAGAGAGAAATAGGATGATGACTTGAATGATGATGATGATGATGATGATGAT 3725
DB 481 CACAGAACAGAGAGAAATAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
OY 3726 ACCTAAAGGAGTACAGCAAGTCCGAAAAAGAGCCATAGGCTTCTGAGATCTGATGAACA 3785
DB 541 ACCTAAAGGAGTACAGCAAGTCCGAAAAAGAGCCATAGGCTTCTGAGATCTGATGAACA 600
OY 3786 GCAGTGGCTGAGAGAAAAAGAGCTCAAGAAGATATATTAAATAATGAAGATGAACAGAA 3845
DB 601 GCAGTGGCTGAGAGAAAAAGAGCTCAAGAAGATATATTAAATAATGAAGATGAACAGAA 660
OY 3846 TAGTCCGCCAAAAAAGGGTAAAGA 3870

DB 661 TAGTCCGCCAAAAAAGGGTAAAGA 665

RESULT 10
CA324097 786 bp mRNA linear EST 26-NOV-2002
LOCUS UI-M-FY0-CCO-m-16-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
DEFINITION IMAGE: 6822425 5', mRNA sequence.
ACCESSION CA324097.1 GI:24542195
VERSION CA324097.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS NIH-MGC
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: egapbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: PYX-5.
FEATURES
source

Location/Qualifiers
1..786
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6822425"
/tissue_type="Whole Brain"
/dev_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NIH_BMAP_FY0"
/note="Organ: Brain; Vector: pYX-Asc; Site:1: Ecor I;
Site:2: Not I; The library was constructed according
to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

BASE COUNT 256 a 171 c 165 g 191 t 3 others
ORIGIN

Query Match 16.1%; Score 671; DB 14; Length 786;
Best Local Similarity 90.7%; Pred. No. 5.3e-105;
Matches 713; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 2794 CGTTTACGCTTCCACTGATATATGCAATCTGTCGCTTGTGCAAAAGATCTCTGA 2853
DB 1 CGTTTACGCTTCCACTGATATATGCAATCTGTCGCTTGTGCAAAAGATCTCTGA 60
OY 2854 AAGGAGAGAGAGCTCATGCTAGGCAATGTTGTGGAATAATTAATGTAAGCGGAG 2913
DB 61 AAGGAGAGAGAGAGCTCATGCTAGGCAATGTTGTGGAATAATTAATGTAAGCGGAG 120

OY	291	TATCTGAAGCAGCAGTGGCTGTAGTAAATAATATATGCTCTCTACAGACATGCT	2973		
Db	121	TACCTGAGCAGCAGTGGCTGTAGTAAATAATATATGCTCTCTACAGACATGCT	180		
OY	2974	GTTCATATACAAATTCACCTTTTGGCAGATGACCAGATATATGTCAAAAGTACAGATATT	3033		
Db	181	GTTCATATATACAAATTCACCTTTTGGCAGATGACCAGATATATGTCAAAAGTACAGATATT	240		
OY	3034	GAAACACTTAAAGATGTTAAAGATGTCTTTGGTTTGTCTCGAATAATTTAATGGCTAAA	3093		
Db	241	GAAACACTTAAAGATGTTAAAGATGTCTTTGGTTTGTCTCGAATAATTTAATGGCTAAA	300		
OY	3094	AATGAAAAATACAGTCAAGCTTTTATATCAGAAAGTGGTAGAATAATTTAAACAAACAAA	3153		
Db	301	AATGAAAAATACAGTCAAGCTTTTATATCAGAAAGTGGTAGAATAATTTAAACAAACAAA	360		
OY	3154	GATGCCCAAGACACAGATGATGCAAAAATGAATGAAAAACTGTACCTGTGTGTGATGTT	3213		
Db	361	GATGCCCAAGACACAGATGATGCAAAAATGAATGAAAAACTGTACCTGTGTGTGATGTT	420		
OY	3214	GCCATGAAATATCATCATGTCTCAAAAGAGTACTACATACATGTTGGAATCTCTAAAGACCG	3273		
Db	421	GCCATGAAATATCATCATGTCTCAAAAGAGTACTACATACATGTTGGAATCTCTAAAGACCG	480		
OY	3274	GTACTACAGCTCGTTTCTTCTACTGCACACCTGACAGAAATTTGAGTACACCAAAATAT	3333		
Db	481	GTTGCTGCAGGCTCGTTTCTTCTACTGCACACCTGACAGAAATTTGAGTACACCAAAATAT	540		
OY	3334	CTGCCTCCCTGAAATGAATTCATTTTCTACTGCTGTGAAAACTTAAACACCAAGTGTCTA	3393		
Db	541	CTGCCTCCCTGAAATGAATTCATTTTCTACTGCTGTGAAAACTTAAACACCAAGTGTCTA	600		
OY	3394	GGAGCTGTTAAACAAGCCACTTTTCATTCACAGGACGACATCTCAGACCAAAATCATCAGA	3453		
Db	601	GGAGCTGTTAAACAAGCCACTTTTCATTCACAGGACGACATCTCAGACCAAAATCATCAGA	660		
OY	3454	ATGGAACATGTAAACGAATGCAAGACGACGCTCAAAATCCAACTCTCCGTGAAGAAATAAG	3513		
Db	661	ATGGAACATGTAAACGAATGCAAGACGACGCTCAAAATCCAACTCTCCGTGAAGAAATAAG	720		
OY	3514	GGGAGGCTTGTAGTATGTTCTGAAATGGATCACAGTGAATAATGAAGATTTACACATGTCTTCA	3573		
Db	721	GGGAGGCTTGTAGTATGTTCTGAAATGGATCACAGTGAATAATGAAGATTTACACATGTCTTCA	780		
OY	3574	CCTTTG 3579			
Db	781	CCTTTG 786			
RESULT 11					
Bg256731					
LOCUS	6032709391	NIH_MGC_92 Homo sapiens	cdna clone IMAGE:4479153 5',		
DEFINITION	mRNA sequence.				
ACCESSION	Bg256731				
VERSION	Bg256731.1	GI:12766547			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1 (bases 1 to 734)				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgabbs-remail.nih.gov				
	Tissue procurement: ATCC				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LMN; at:				

FEATURES	source	1..734
ORGANISM	source	/organism="Homo sapiens"
MOLECULAR	source	/mol_type="mRNA"
DB	source	/db_xref="taxon:9606"
CLONE	source	/clone="IMAGE:4479153"
TISSUE	source	/tissue_type="embryonal carcinoma, cell line"
LAB	source	/lab_host="DH10B (phage-resistant)"
NOTE	source	/note="Organ: testis; Vector: PCWV-SPORT6; Site:1: NotI; Site:2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."
BASE COUNT	278 a	154 c 161 g 141 t
ORIGIN		
Query Match	15.9%;	Score 663.4; DB 10; Length 734;
Best Local Similarity	96.6%;	Pred. No. 1.1e-103;
Matches 700; Conservative	0;	Mismatches 21; Indels 4; Gaps 2;
QY	3322	TCGAAGACTACTACATACAGTTTGGAAATCTCTTAAGACCCGGTACTACACAGCTGTTTC 3291
DB	1	TCAAAAGACTACATACATACAG-TTGGAAATCTCTTAAGACCCGGTACTACACAGCTGTTTC 59
QY	3292	TTCACTCACTGACCAAGAAATTTGAGTACACCAAAAATTTCTGCTCCCTGAATGAAA 3351
DB	60	TTCACTCACTGACCAAGAAATTTGAGTACACCAAAAATTTCTGCTCCCTGAATGAAA 119
QY	3352	TCATTTTTCACCTCTGGAACCACTAACAACAACATGTTCTGAGAGCTGTTTACAGCCA 3411
DB	120	TCATTTTTCACCTCTGGAACCACTAACAACAACATGTTCTGAGAGCTGTTTACAGCCA 179
QY	3412	CTTTCATCAGAGCAACCAATCTCAGCCCAATCATCAGCAATGGAAGCAACGTGAAGAT 3471
DB	180	CTTTCATCAGAGCAACCAATCTCAGCCCAATCATCAGCAATGGAAGCAACGTGAAGAT 239
QY	3472	GCAAGCAGACACTCAAAATCCCAAGCTCTCTGGAAGATTAAGGGGAGGCTTGATAGTCT 3531
DB	240	GCAAGCAGACACTCAAAATCCCAAGCTCTCTGGAAGATTAAGGGGAGGCTTGATAGTCT 299
QY	3532	GAAATGATCATCAGTGAATTAAGATTACCAATGTCTTACCTTTGCCGGGAAAAAAA 3591
DB	300	GAAATGATCATCAGTGAATTAAGATTACCAATGTCTTACCTTTGCCGGGAAAAAAA 359
QY	3592	AGTGACAAGACAGACACTCTGATCTTGTAAAGTCTGATGATGGAAGACCTAGAGGACG 3651
DB	360	AGTGACAAGACAGACACTCTGATCTTGTAAAGTCTGATGATGGAAGACCTAGAGGACG 419
QY	3652	AAAAAAGCGCCGCTACAGAAACAGAGGAGAAATTTAGTATGATGACTGACTGATG 3711
DB	420	AAAAAAGCGCCGCTACAGAAACAGAGGAGAAATTTAGTATGATGACTGACTGATG 479
QY	3712	GTAACAGGAACGAAACCTTAAGAGGAGTACAGGAAGTGGAAAAAGGCAATAGGCTTCA 3771
DB	480	GTAACAGGAACGAAACCTTAAGAGGAGTACAGGAAGTGGAAAAAGGCAATAGGCTTCA 539
QY	3772	GAATCTGATGAAACAGCACTGGCTGAGAGAAAAAGGCTCAAGAAGATATATTAGAAAA- 3830
DB	540	GAATCTGATGAAACAGCACTGGCTGAGAGAAAAAGGCTCAAGAAGATATATTAGAAAA- 599
QY	3831	--TGAAGTGAACAGAAATAGTCCGCCCAAAAAAGGTTAAAGAGGCCCAACCAAAAACCT 3888
DB	600	CTGAACGATGAAACAGAAATAGTCCGCCCAAAAAAGGTTAAAGAGGCCCAACCAAAAACCT 659
QY	3889	CTGGTGAGGATACACCAAAAGAGGACCAATGAAATCTTCTAAAGAAAGAGCAAA 3948
DB	660	CTGGTGAGGATACACCAAAAGAGGACCAATGAAATCTTCTATTAAGAGAGCAAA 719
QY	3949	AAAAA 3953

Db 720 CAAA 724

RESULT 12
CB312535
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CB312535 771 bp mRNA linear EST 04-MAR-2003
AGENCOURT_11835698 NICHD_Rh_Ov1 Macaca mulatta cDNA clone
IMAGE:6892555 5', mRNA sequence.
CB312535
CB312535.1 GI:28835251
EST.
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (bases 1 to 771)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Eliot Spindel
cDNA Library Preparation: CLOUTECH
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINT at:
<http://image.llnl.gov>
Plate: LICM3155 row: 9 column: 18
High quality sequence stop: 563.
Location/Qualifiers
1..771

FEATURES
source
/organism="Macaca mulatta"
/mol_type="mRNA"
/db_xref="taxon:9544"
/clone="IMAGE:6892555"
/tissue_type="Ovary"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_Rh_Ov1"
/note="Organ: ovary; Vector: pDNR-LIB; Site1: Sfi I;
Site2: Sfi I; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.0-4.0 kb. Tissue pooled from
pre-pubertal, post pubertal sn menopausal monkeys.
Constructed by Clontech. Note: this is a NICHD library."

BASE COUNT 291 a 144 c 159 g 177 t

ORIGIN

Query Match 15.8%; Score 660.6; DB 14; Length 771;
Best Local Similarity 96.0%; Pred. No. 3.2e-103;
Matches 744; Conservative 0; Mismatches 19; Indels 12; Gaps 6;

3014 ATGTCAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGTTGGTTGTC 3073
|||||
3 ATGTCAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGTTGGTTGTC 62
|||||
3074 TGGAAATATTAAATGCTTAATAATGAATTAACAGTCAGCGCTTTTATCAGAAAGATGTA 3133
|||||
63 TGGAAATATTAAATGCTTAATAATGAATTAACAGTCAGTCATTTATCAGAAAGATGTA 122
|||||
3134 AAAATATTAAACAACAAGATGCCCAAGACCGATGATGCAAAAATGAATGAAGAAC 3193
|||||
123 AAAATATTAAACAACAAGATGCCCAAGACCGATGATGCAAAAATGAAGAAC 182
|||||
3194 TGTACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3253
|||||
183 TGTATACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 242
|||||
3254 TGGATCTCTCTAAAGACCGGACTACGACGCTGTTTCTTCACTCACTGACAGAAATT 3313
|||||
243 TGGATCTCTCTAAAGACCGGACTACGACGCTGTTTCTTCACTCACTGACAGAAATT 302

QY 3314 TCAGTACACCAAAATTTATCTGCGCTCGAAGTAAATCATTTTCACTCCTGGAAAC 3373
|||||
Db 303 TCAGTACACCAAAATTTATCTGCGCTCGAAGTAAATCATTTTCACTCCTGGAAAC 362
|||||
QY 3374 CTAAACAACCAATGTTCTAGAGAGCTGTATACAAAGCACTTTCATCAGAGCAAGCAT 3433
|||||
Db 363 CTAAACAACCAATGTTCTAGAGAGCTGTATACAAAGCACTTTCATCAGAGCAAGCAT 422
|||||
QY 3434 CTCAGACCAATTCATCAGATGGAATGGAATCTGTAAAGCAATGCAAGCAGCTCAATCCAA 3493
|||||
Db 423 CTCAGACCAATTCATCAGATGGAATGGAATCTGTAAAGCAATGCAAGCAGCTCAATCCAA 482
|||||
QY 3494 GCTCTCCTGGAAGAAATAAAGGGAGGCTGTAGATTCGAAATGATCAGTGAATG 3553
|||||
Db 483 GCTCTCCTGGAAGAAATAAAGGGAGGCTGTAGATTCGAAATGATCAGTGAATG 542
|||||
QY 3554 AAGATTACCAATGCTCTCACCCTTTCGCGGGAAAAAAGTACCAAGAGAGAGACTCTG 3613
|||||
Db 543 AAGATTACCAATGCTCTCACCCTTTCGCGGGAAAAAAGTACCAAGAGAGAGACTCTG 602
|||||
QY 3614 ATCTTGAAGTCTGATTTGAGAGAGCCCTAGAGGCAAGAAAAAGCCCTCAGAGAAC 3673
|||||
Db 603 ATCTT-----TCTGAATTGAGAGAGCCCTAGAGGCAAGAAAAAGCCCTCAGAGAAC 656
|||||
QY 3674 AGGAG-GAGAAATTTGGTATGATGAC-TTGACTAAGTTGGTACAGAGAAC-AGAAACCTA 3730
|||||
Db 657 AGGAGGAGAAATTTGGTATGATGACTTTGACTAAGTTGGTACAGAGAACAGAAACCTA 716
|||||
QY 3731 AAGGAGTCTAG-CGAAGTCGAAAGA--GGCCATACGGCTTCAGATCTGATGA 3782
|||||
Db 717 AAGGAGTCTAGCGGAGGTGGAAGAAAGGACATACGGCTTCGATCTGATGA 771
|||||

RESULT 13
CA405800
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CA405800 695 bp mRNA linear EST 07-NOV-2002
1001949 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
CDNA 5', mRNA sequence.
CA405800
CA405800.1 GI:24770671
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 695)
Yang, R.-Z., Shuldiner, A. and Gong, D.-W.
EST analysis of human adipose gene expression
unpublished
Contact: Gong Da-Wei
Division of Endocrinology, Diabetes and Nutrition
University of Maryland
660 Redwood St., HH497, Baltimore, MD 21201, USA
Tel: 410 706 1672
Fax: 410 706 1622
Email: dgong@medicine.umaryland.edu
PCR Primers
FORWARD: CTCGGGAAGCGCGCCATGTTGGT
BACKWARD: AATAGACTACTATAGGCGCAATTGG
Seq primer: GTTGGTACCGGGAATTC.
Location/Qualifiers
1..695

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="Male and Female"
/tissue_type="Adipose"
/clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
/note="Vector: lambda triplex"

BASE COUNT 222 a 154 c 128 g 191 t

ORIGIN

Query Match 15.7%; Score 657.2; DB 14; Length 695;
 Best Local Similarity 99.1%; Pred. No. 1.2e-102;
 Matches 692; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 1680 AAGCGATTGAGTACTTGTAGTCCACATGCTCCGTGACAGGCTGAAGTTGTG 1739
 DB 1 AAGCGATTGAGTACTTGTAGTCCACATGCTCCGTGACAGGCTGAAGTTGTG 60
 QY 1740 GCGTGAATTAAGTAAAGTTGGCAACCCCAAGACGCTTCAAAATCCTTCTGGAAT 1799
 DB 61 GCGTGAATTAAGTAAAGTTGGCAACCCCAAGACGCTTCAAAATCCTTCTGGAAT 120
 QY 1800 GATCAAGTTCTCTTGGAGAGATAGACCTGTGCACATGATACGGAATCTATCACTGC 1859
 DB 121 GATCAAGTTCTCTTGGAGAGATAGACCTGTGCACATGATACGGAATCTATCACTGC 180
 QY 1860 TCTTATTAACAAGTGAACAATCAATAGATGAACAGCATGATGAATGAGGTGT 1919
 DB 181 TCTTATTAACAAGTGAACAATCAATAGATGAACAGCATGATGAATGAGGTGT 240
 QY 1920 TCCAACTGATCAAGCCATCAGACAGGCTTGAATGCTTAAAGTACTTCAATTACACA 1979
 DB 241 TCCAACTGATCAAGCCATCAGACAGGCTTGAATGCTTAAAGTACTTCAATTACACA 300
 QY 1980 TCCCATCTCATTTCAATCTGCTGAACATTTGAATCATCTGCTGCTGAACATTTGA 2039
 DB 301 TCCCATCTCATTTCAATCTGCTGAACATTTGAATCATCTGCTGCTGAACATTTGA 360
 QY 2040 TGATGAAAAAGTAGACAGACAGCTCAGTACAAATTTTCAAAAACAGAGAAACAATTTGA 2099
 DB 361 TGATGAAAAAGTAGACAGACAGCTCAGTACAAATTTTCAAAAACAGAGAAACAATTTGA 420
 QY 2100 AGAGGATTTTCCACACATCAGATCAGCTTGTCTTCTTTCATCATCAAAATCTAAAA 2159
 DB 421 AGAGGATTTTCCACACATCAGATCAGCTTGTCTTCTTTCATCATCAAAATCTAAAA 480
 QY 2160 AGAGACCCCGCCGTCACAAATATGCAATTCATTCATTCATTCATTCATTCATTCATTA 2219
 DB 481 AGAGACCCCGCCGTCACAAATATGCAATTCATTCATTCATTCATTCATTCATTCATTA 540
 QY 2220 AGAGACCCAGTTTGCACAGATATTTGAGCCTGTGCATTAAGAGCTAGATCCAAACACT 2279
 DB 541 AGAGACCCAGTTTGCACAGATATTTGAGCCTGTGCATTAAGAGCTAGATCCAAACACT 600
 QY 2280 GGAACATCTCATTAACACATTTGTTACTATTTGTCATATTTGTCCTTGCACCTGATCA 2339
 DB 601 GGAACATCTCATTAACACATTTGTTACTATTTGTCATATTTGTCCTTGCACCTGATCA 658
 QY 2340 ATTGGCTGCTTGGAAATCTGGTGAAGTACTTCA 2377
 DB 659 ATTGGCTG-TCCTTGAATCTTGGTGAAGTACTTCA 695

RESULT 14
 BU703488 786 bp mRNA linear EST 09-OCT-2002
 LOCUS BU703488
 DEFINITION UI-M-F00-b20-g-20-0-UI.r1 NIH-BMAP_F00 Mus musculus cDNA clone
 IMAGE: 6405163 5', mRNA sequence.
 ACCESSION BU703488
 VERSION BU703488.1 GI:23629371
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 786)
 TITLE NIH-MGC http://mgc.nci.nih.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pyx-5.
 location/Qualifiers
 1. 786
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6405163"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone.lib="NIH-BMAP_F00"
 /note="Organ: Brain; Vector: pyx-Asc; Site_1: EcoR I;
 Site_2: Not I; The library was constructed according
 to the protocol of Bontalio, Lennan and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction. ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TGAGAGAGCC. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
 program coordinator."

BASE COUNT 248 a 143 c 152 g 242 t 1 others
 ORIGIN

Query Match 15.7%; Score 653.8; DB 13; Length 786;
 Best Local Similarity 91.1%; Pred. No. 4.7e-102;
 Matches 694; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 19 AGACCAATGATGAAAAATTAACATATCCCTGGGGTCAAGAAATATCAGATTAATA 78
 DB 25 AGACCAATGATGAAAAATTAACATATCCCTGGGGTCAAGAAATATCAGATTAATA 84
 QY 79 TCTAAGAGAGATGTGAGACGATTAAGATGGTGTGAACCTTTATGATATGAGAC 138
 DB 85 TCTAAGAGAGATGTGAGACGATTAAGATGGTGTGAACCTTTATGATATGAGAC 144
 QY 139 CAGACCTCTAAGAAAGAGAGCTTATTAACTTACCTTACCTTCTGCTAGAT 198
 DB 145 CAGACCTCTAAGAAAGAGAGCTTATTAACTTACCTTACCTTCTGCTAGAT 204
 QY 199 TTTTCTCAAGCATCCCTGGTAAAGATGTTCGCTTACTGTAGCTCCTCTGCTGAT 258
 DB 205 TTTTCTCAAGCATCCCTGGTAAAGATGTTCGCTTACTGTAGCTCCTCTGCTGAT 264
 QY 259 ATTTTCAGAGTTTATGCTCCTGGAAGCTCCTTACACATCCCTGATTAAGATTAAGATATA 318
 DB 265 ATTTTCAGAGTTTATGCTCCTGGAAGCTCCTTACACATCCCTGATTAAGATTAAGATATA 324
 QY 319 TTTATGTTTATTAACAACAGCTTGAAGGGCTAGAGATCAAAAGAGCCCAATTAAT 378
 DB 325 TTTATGTTTATTAACAACAGCTTGAAGGGCTAGAGATCAAAAGAGCCCAATTAAT 384
 QY 379 AGGTATTTTATTTACTTGAACATGCTGGGTCAAGCATATATACATATATGCTTGA 438
 DB 385 AGGTATTTTATTTACTTGAACATGCTGGGTCAAGCATATATACATATATGCTTGA 444
 QY 439 TTGAAGATAGCATGAATTTTCCACCAGCTATACGAACCTTATTTGAGTTATTAAC 498
 DB 445 TTGAAGATAGCATGAATTTTCTTACCACTTACAGAACATATTTCTCAGTTATTAAC 504

QY 499 AATGGCCAAATCAGAAAGTCCATATGACAGTGTAGACCTTATGAGCTCTATATTGCT 558
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 505 AATGGCCAAATCAGAAAGTCCATATGACAGTGTAGACCTTATGAGCTCTATATTGCT 564
 QY 559 GAAGGTGATACAGTGTCTCAGAGCTTTTGGATACGCTTTTAAATCTGTACCTGCT 618
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 565 GAAGGTGATACAGTGTCTCAGAGCTTTTGGATACGCTTTTAAATCTGTACCTGCT 624
 QY 619 CATTAAATTTAAACAGAGCATATGATTTGGCAAGGCTTTACTGAAAGGACAGCT 678
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 625 CATTAAATTTAAACAGAGCATATGATTTGGCAAGGCTTTACTGAAAGGACAGCT 684
 QY 679 CAAGCTATTTGAGCCATATATATACCACTTTTAAATCAGGCTCTGATGCTGGGAAACA 738
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 685 CAAGCTATTTGAGCCATATATATACCACTTTTAAATCAGGCTCTGATGCTGGGAAACA 744
 QY 739 TCTATCAGGATTTGTCTCAGAGCATGCTTTGACTTAATTTTG 780
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 745 TCTATCAGGATTTGTCTCAGAGCATGCTTTGACTTAATTTTG 786

RESULT 15

CA324284

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA324284 779 bp mRNA linear EST 26-NOV-2002
 UI-M-FY0-ccp-a-07-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
 IMAGE: 6822512 5', mRNA sequence.
 CA324284
 EST: 24542382
 Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 779)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.

FEATURES

source

Location/Qualifiers
 1..779
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6822512"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpf"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_FY0"
 /note="Organ: Brain; Vector: pyx-Asc; Site: 1: EcoR I;
 Site: 2: Not I; The library was constructed according
 to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is ACCGACAG. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH), Hsin-Chin, Ph.D.,
 program coordinator."

BASE COUNT 241 a 166 c 174 g 197 t 1 others

Query Match

Best Local Similarity

Matches 699: Conservative 0; Mismatches 80; Indels 1; Gaps 1;

15.4%; Score 640.6; DB 14; Length 779;

89.6%; Pred. No. 8.6e-100;

1 GCGATTTGACAGAAACAGGAAATTAAGTAACAGATATGTCAGCCCTGAGACCTGCG 60

2585 GAGACTTGACAGACAGGAGGAAATTAAGTAACAGATATGTCAGCTGTGCTG 2644

1 GCGATTTGACAGAAACAGGAAATTAAGTAACAGATATGTCAGCCCTGAGACCTGCG 60

2645 CTGGAGAGCTATTTGTAAGCTGGACAGAAACCTGTTACCATGAAATCAGACATTTG 2704

61 CTGGAGAGCTATTTGTAAGCTGGACAGAAACCTGTTACCATGAAATCAGACATTTG 120

2705 AACCAATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2764

121 AGCAGTACAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

2765 TTGCCAGAACTTACAAAGGCTTTCCGTTTACGCTTCCACTGATATATGCA 2824

181 TCGCTCAGAACTTACAAAGGCTTTCCGTTTACGCTTCCACTGATATATGCA 240

2825 TCTGTGCTTTGTCGCAAAAGATTCCTGTAAGGAGAGAGAGAGAGAGAGAGAGAGAG 2884

241 TCTGTGCTTTGTCGCAAAAGATTCCTGTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

2885 TGGTGAAGAAATTAATGTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2944

301 TGGTGAAGAAATTAATGTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

2945 AATTATGCTCTTCTACAGAGATGTTGTTCCATATGATGATGATGATGATGATGATGAT 3004

361 AATTATGCTCTTCTACAGAGATGTTGTTCCATATGATGATGATGATGATGATGATGAT 420

3005 ACCGATATATGTCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3064

421 ACCGATATATGTCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

3065 GCTTGTCTGAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3124

481 GCTTGTCTGAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540

3125 AGATGTAGAAATTAATTAACAAACAAAGATGCTCCAGAGACAGATGATGATGATGATGAT 3184

541 AATGTAGAAATTAATTAACAAACAAAGATGCTCCAGAGACAGATGATGATGATGATGATGAT 600

3185 ATGAAACTGTACAGCTGTGTGATGTTGCCATGATATGATGATGATGATGATGATGATGAT 3244

601 ATGAAACTGTACAGCTGTGTGATGTTGCCATGATATGATGATGATGATGATGATGATGAT 660

3245 CATACAGTTTGAATCTCTTAAGACCCGGTACTACAGCTGTCTTCTACACTCAACCTG 3304

661 CGTACAGCTGTGAGCTCTCTTAAGACCCCGTCTGAGCTGTCTTCTACACTCAACCTG 720

3305 ACAAGATTTTACAGTACACCAAAATTAATCTGCTCTGAAATGAATGAATGAATGAATGAAT 3364

721 ACAAGATTTTACAGTACACCAAAATTAATCTGCTCTGAAATGAATGAATGAATGAATGAAT 779

Search completed: September 24, 2003, 16:28:07
 Job time : 8082 secs

THIS PAGE BLANK (USPTO)

QY 2507 TACTGGAATGAAAAATATACAGTAATACAGAACTTCTACCTTAAGATTGCTAACAA 2566
 Db 203 TGTGGGTATGAAAAACACAGCTATAATCTGCCAATTCACACCTTGCTGGTTATATACAG 262
 QY 2567 CAATATTCATATGATGATGAGAACTTGACAGAAAGGAGAAAAATTAATTAACCAATANGT 2626
 Db 263 CGATGTTGGTTAGTAGAGGTTGCTGACAGAGCAAAAAGAGATCACTGTAATCTGTAATAGT 322
 QY 2627 CACGTCTGAGACTTGCTGGAGTGTCTATTTGTAAGCTGGACACAAAGACCTGTTACC 2686
 Db 323 CTCCTCTCCGATTAAGTCTGCTGCTAGTCCATATGAGACTTCTGCTAGAACCTTGTACC 382
 QY 2687 ATGAAATCATCATTAAGAACAAATATCAGTATGCTATTAATCAACATGATGATCT 2746
 Db 383 ATGAAATTAATTAACCCACAGACAGTTTCAGCTCTGTCAGCTTGTATTAATGATGAGTGT 442
 QY 2747 ATCAAGTAAGCAAGTGTGTTCCAGAACTTCACAAAGGCTTCCGCTTACGGCTTC 2806
 Db 443 ACCAAGTAAGCAATATTTGCTCAGAAAGCTGCAATAAAGCACTTGTGAAGTACTGCTCC 502
 QY 2807 CACTGAGTATATGCAATCTGTGCTCCCTTGTGCAAAAGATCCTGTAAGAGAGAGAGAG 2866
 Db 503 CATGAGATATATGCGGATCTTGTGCTGTGTCGCAAAAGATCCTGTGAAGAGAGAGAGAG 562
 QY 2867 CTCAATGCTAGCAATGTTTGTGTAATAATTAATTAAGCGGAGATATCTGAAGCAGC 2926
 Db 563 CACAGCGACGACAAATGTTTACTGAAAAATATCAGATACGACAGGGAATATCAATTAAGCAGA 622
 QY 2927 ATGACAGCTGTAGAAAAATTAATGCTCTTACACAGAGATGCTGCTTCCATATACAA 2986
 Db 623 ATCTAATGCTACTGAGAAATTAATTAATCTGCTGCTGTAATGATGATCTCATACAGTA 682
 QY 2987 TTCACCTTTTGGACATGACCCAGATTAATGTCAAAGTACAGAGATTAAGCAACTTAAG 3046
 Db 683 TTCACCTCTGAGCCCATGATCCAGATTTTACAAGATCAAGAGATGATCAGCTGCTG 742
 QY 3047 ATGTTAAAGATGTCCTTGTGTTGTTGCTGGAATTTATGCTTAATAATGAATAACAA 3106
 Db 743 ATATCAAAAGATGCTGCTGATGCTTGAAGTTTAAATGACAAAGATGAAAAACAATA 802
 QY 3107 GTCACGCTTTTATCAGAAAGATGAGTAAGTAATTAACAAACAAAGATGCCCCAGAGAC 3166
 Db 803 GCCATGCTTTATGAGAGAGATGCGACAGAAACATCAAGTTAACAGAGATGCCAGTCTC 862
 QY 3167 CAGATGATGCAAAAAATGAAATGAAAACTGACACTGTGTGATGTTGCCATGATATCA 3226
 Db 863 CAGATGAATCCAGCAAAATGAAAACTGTATACAGTATGATGATGCTCTCTGTGTTA 922
 QY 3227 TCATGTCAAAGATGATCATACAGTTTGAATCTCTTAAGACCCGCTACTACAGCTC 3286
 Db 923 TAAATAGTAAAGTGGCTTTGTGCA - ATGCAAGATTCACCAAGGACCCGCTCCAAATGAA 980
 QY 3287 GTTCTCTACCTCAACCTGACAAATTTCAATACACCAAAAATTAATGCTGCTCTCTGAAA 3346
 Db 981 ATTTTACAC - ACCTGAAAAGAGCTTCTGTATAGATTAAGATTAATTTTCAGAGAGAA 1037
 QY 3347 TGAATCATTTTTCACCTCTCGTGAAGAACTTAACCAACCAATGTTCTAGAGCTGTTACA 3406
 Db 1038 CAAAGATGATCTCTGTAAAGAGAAAGCCGCTGAGAGTACTAGTGCAGTAATAA 1097
 QY 3407 AGCCACTTTCATGACGAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAAGTGA 3466
 Db 1098 AGCCTTATACGCAACGGGAAGAAACCTTATGTTAGAGCACTGCGACTAGACTGGAA 1157
 QY 3467 GCATGCAAGCAGAGCTCAAAATCCAACTCTCTCGAAGAAATTAAGGAGGCTGTGTA 3526
 Db 1158 GCATTAATTAATTAATTCAGAGCTGAACCTTCAACCGAAATTCATCAAGGGAACAGA 1217
 QY 3527 GTTCTGA 3533
 Db 1218 GTTCAGA 1224

RESULT 2
 US-10-040-739-107
 ; Sequence 107, Application US/10040739
 ; Publication No. US20020173635A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; McCoy, John
 ; LaVallie, Edward
 ; Racie, Lisa
 ; Merberg, David
 ; Teacy, Maurice
 ; Spaulding, Vikki
 ; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
 ; NUMBER OF SEQUENCES: 1519
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/040,739
 ; FILING DATE: 07-Jan-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/036,520
 ; FILING DATE: 03-JUN-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brown, Scott A.
 ; REGISTRATION NUMBER: 32,724
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 498-8224
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 439 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 107:
 US-10-040-739-107
 Query Match 9.9%; Score 411.4; DB 13; Length 439;
 Best Local Similarity 98.6%; Pred. No. 1.5e-91;
 Matches 415; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 3216 CATGAATATCATATGATGTAAGAGATGATACATGATGTAATCTCTTAAGACCCGCT 3275
 Db 13 CATGCCCTACATCATCATGCTCAAGAGATGATACATGATGTAATCTCTTAAGACCCGCT 72
 QY 3276 ACTACACGCTGCTTCTTCTACCTCAACCTGACAGCAATTTCAAGTAAACCAAAATTAATCT 3335
 Db 73 ACTACACGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 132
 QY 3336 GCTCTCGTAATGAATCATTTTTCACCTCTGGAAGAACTTAACCAACCAATGTTCTAGG 3395
 Db 133 GCTCTCGTAATGAATCATTTTTCACCTCTGGAAGAACTTAACCAACCAATGTTCTAGG 192
 QY 3396 AGCTGTAAACAGCCACTTTATCAGAGGAGCAAGCAATCTCAGACCAATATCAGCAAT 3455
 Db 193 AGCTGTAAACAGCCACTTTATCAGAGGAGCAAGCAATCTCAGACCAATATCAGCAAT 252
 QY 3456 GGAACCTTAAGCAATGCAACAGCAGCTCAAAATCCAAGCTCTCTGGAAGAAATTAAGG 3515
 Db 253 GGAACCTTAAGCAATGCAACAGCAGCTCAAAATCCAAGCTCTCTGGAAGAAATTAAGG 312

Query	Match	5.2%	Score 217.6;	DB 11;	Length 471;
3253	TTGGAATCTCTCTAAAGACCGGGTCTACACCTGTTTCTTCATCTCAACGTCAGAT	69.08;	Pred. No. 1.8e-43;		
362	GCACATTCACCAAGAGACCAGTCTCCCAATGAATTTTATACACAACGTAAGAGAC	0;	Mismatches 134;	Indels 0;	Gaps 0;
3313	TTTCAGTAACACCAAAAATTTATCTCCCTCTGTAATGAATCATTTTCTACTCTCGAAA				
422	TTCTGTAAACGTAAGATGATTTATTTTCAGAAAGACAAAGATGATCTGTTAAAGAAAG				
3373	CCCTAAACACCAATGTTTCTAGAGAGCTGTTTAAACAGCCACTTTATCATCAGACGCAACAA				
482	CCAAAGCTCTGTGAGTACTGAGTGCAGTAATTAAGCTTTATTCAGCAACGCGAAAGAAA				
3433	TTCTAGACCAAAATATATACAGAAATGGAACGTAAAGCAATGCAACGACAGCTCAATCCA				
542	CCCTATGTTTGAAGACCTCGGACCTGAGACCTGGAAGCAATATTTATGTAATTTCAAGACTG				
3493	AGCTCTCCCTGGAAGATTAAGGGAGGAGCTTGATAGTTCTGA 3533				
602	AACCTTCACACCGGAATTCATCAAGGACAGAGTTTCAGA 642				
US-09-918-995-3514/c					
Sequence 3514, Application US/09918995					
Publication No. US20030073623A1					
GENERAL INFORMATION:					
APPLICANT: Hyseq, Inc.					
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED					
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES					
FILE REFERENCE: 20411-756					
CURRENT APPLICATION NUMBER: US/09/918,995					
CURRENT FILING DATE: 2001-07-30					
PRIOR APPLICATION NUMBER: US/09/235,076					
PRIOR FILING DATE: 1999-01-20					
NUMBER OF SEQ ID NOS: 38054					
SOFTWARE: FastSeq for Windows Version 3.0					
SEQ ID NO 3514					
LENGTH: 471					
TYPE: DNA					
ORGANISM: Homo sapiens					
US-09-918-995-3514					
Query Match					
Best Local Similarity	69.08;				
Matches 298; Conservative	0;				
1309	AAAGACAAATTTGCATATATATATATATCAAAATAGTATGATGATGCACTACTTGTGAA				
459	AAAGGCAACTCTTCGATATTTATATACAGACAGATTTGACGACAAACTGTTGTAAG				
1369	CGGATCTTGGCTCAATACATGTTCTGCACATTTAGAAACTACAGAACGAGTAATGC				
399	AAATCTTTCCTCGATCTTCTTGTCCCAACCTGGAAACAGAAAGAGAAATGAATGC				
1429	TTATATACTTGTATGTCACACAGATTTTAATGCTCTGAAAGCATGTAATGAATGTG				
339	TTATATACTTGTATGTCACACAGATTTTAATGCTCTGAAAGCATGTAATGAATGTG				
1489	AAATGTCAAATATGCTCTCCACATCAATCAATGAAGATTTGCTTGACTGATTAAAGCAACC				
279	AAGTGTCAAGAACTGCTTGGAGGACATGACGGAACATTAATGATTTTGCAACAGCACT				
1549	AAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTATGTTTATTAACAGAAATTTA				
219	ACATCAAGAGCTAACTGTTCTGCTATGTTTGAAAACTGATGACCATAGCAAAAGAAATTTG				
1609	CCGATCTCTGTAAGGCTCAGATTTTCATAGAAATTTCAACAGAGTGTAGAGATGAT				
159	CTGACCCCGGGAAGACAAAGATTTTGTGAGAAATTTTAAACAGAGTTCTTCGGCGATGAT				
1669	GAGAAATAAGAAAGCATTAAGATGACTTGTATAGTCAACATGCTCTGCAAGCAGCT				

DB 99 GAGAACTTCGCTCACTTGAGTATTATTAATAGCCCAACCTGTTTCGCAACAGACA 40
QY 1729 GAAGTTGTGTG 1740
DB 39 GATATTTGTGTG 28

RESULT 5

US-09-879-536-682/c
; Sequence 682, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; TITLE OF INVENTION: PRODUCTS

; FILE REFERENCE: CCD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/879,536

; CURRENT FILING DATE: 2001-09-21

; PRIOR APPLICATION NUMBER: US 60/088,801

; PRIOR FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 682

; LENGTH: 530

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-879-536-682

Query Match
Best Local Similarity 98.1%; Score 203.6; DB 10; Length 530;
Pred. No. 5.7e-40;
Matches 206; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 104 TAAAGATGTTGTGAAAACTTTATGATATGAGCAGCAGACTGTGAGAGAAAAAGAGC 163
DB 218 TATGATGCTGTGTAACCTTTATGATATGAGCAGCAGACTGTGAGAGAAAAAGAGC 159
QY 164 TTTATTAACTGCTTACATCTTGCTCAGATTTTCTCAAGATCCTGTGAAG 223
DB 158 TTTATTAACTGCTTACATCTTGCTCAGATTTTCTCAAGATCCTGTGAAG 99
QY 224 ATGTTGCTTACTGTAGCCTGCTGCTTGTGATATTTTCAGATTTATGCTCCTGAAG 283
DB 98 ATGTTGCTTACTGTAGCCTGCTGCTTGTGATATTTTCAGATTTATGCTCCTGAAG 39
QY 284 CTCCTTACACATCCCTGATTAAG 313
DB 38 CTCCTTACACATCCCTGATTAAG 9

RESULT 6

US-09-918-995-33817

; Sequence 33817, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/735,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33817
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33817

Query Match
Best Local Similarity 100.0%; Score 164; DB 11; Length 433;
Pred. No. 3.3e-30;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4009 ACGGACGAGAGTCCAAAGCAACGACCGAGTGTCAAGAGACGACAGAGAGCA 4068
DB 3 ACGGAAACGAGAGTCCAAAGCAACGACCGAGTGTCAAGAGACGACAGAGAGCA 62
QY 4069 GAATCTCGAATCTAGTCAATTAATCAACGACGAGTGTCAAGAGAGAGAGAGCA 4128
DB 63 GAATCTCGAATCTAGTCAATTAATCAACGACGAGTGTCAAGAGAGAGAGAGCA 122
QY 4129 AGACCATCAAAAACGCCATCACCATCAGACCAAAAAAATGT 4172
DB 123 AGACCATCAAAAACGCCATCACCATCAGACCAAAAAAATGT 166

RESULT 7

US-09-918-995-20720

; Sequence 20720, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/735,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 20720

; LENGTH: 413

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-918-995-20720

Query Match
Best Local Similarity 72.3%; Score 96.4; DB 11; Length 413;
Pred. No. 1.8e-13;
Matches 121; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 107 AGATGTTGTGAAAACTTTATGATATGAGCAGCAGACTGTGAGAGAAAAAGAGCTTT 166
DB 252 AGATGTTGTGAAAACTTTATGATATGAGCAGCAGACTGTGAGAGAAAAAGAGCTTT 311
QY 167 ATTAAACCTAGCTTTACATCTTGCTTCAAGATTTTCTCAAGCATCTGTAAGATG 226
DB 312 ATCTCCACATAGCCTTGCATCTTGATCTGAATCTTCTCAGGAACCCCAATAAGATG 371
QY 227 TTCGCTTACTGTAGCCTGCTGCTTGCATCTTGATCTTGAGATTTTCAGGA 268
DB 372 TCGCTTCTCTGTAGCATGTGTGTGGCTGATATCTTCGTA 413

RESULT 8

US-09-728-445-172

; Sequence 172, Application US/09728445

; Patent No. US20020102543A1

; GENERAL INFORMATION:

; APPLICANT: Friedlich, Glenn

; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and

; TITLE OF INVENTION: Animals


```

; FEATURE:
; NAME/KEY: unsure
; LOCATION: 8289, 8310, 8313
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1931

```

Query Match	1.5%;	Score 64.4;	DB 12;	Length 14006;
Best Local Similarity	48.4%;	Pred. No. 0.00015;		
Matches 179; Conservative	0;	Mismatches 191;	Indels 0;	Gaps 0;

QY	3800	AAAAGAGGCTCAAAAGAGATTTATTAGAAAATGAAGATGAACAGATATGTCGCCCAAAA	3859
Db	2484	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2425
QY	3860	AGGGTAAAGAGGCCGACCAACCAAAACCTCTTGGTGAGGTACACAAAAGAAAGCCAA	3919
Db	2424	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTATTAATAAAAAA	2365
QY	3920	CAATGAAACCTTTAAAAAAAAGGAAGCAAAAAAAAAATCTGCACCTCCAGCACAGAGAGG	3979
Db	2364	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTAAAAAATAAAAAAAAAAAAA	2305
QY	3980	AGGAAGAAAGAAAGACAAAGTGGAAATACGAAACGAAAGTCCAAAAAGCAACAGCAC	4039
Db	2304	AACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAGC	2245
QY	4040	GAGTGTCAAGAGAGCACAGACAGACAGAAATCTCGAATCTAATGCAATGATTCCA	4099
Db	2244	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAA	2185
QY	4100	CACAGTCCACACACAGAAAGGACAGAGAAACACATCAAAAAACCCATCACATCACAAAC	4159
Db	2184	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCAAAAAAAAAAAAAAAAAAAAA	2125
QY	4160	CAAAAAAAAA 4169	
Db	2124	AAAAAAAAAA 2115	

```

RESULT 12
US-10-311-455-1670/c
: Sequence 1670, Application US/10311455
: Publication No. US20030143606A1
: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determin
: FILE REFERENCE: 5013.1014
: CURRENT APPLICATION NUMBER: US/10/311.455
: CURRENT FILING DATE: 2002-12-16
: PRIOR APPLICATION NUMBER: PCT/EP01/07537
: PRIOR FILING DATE: 2001-07-02
: PRIOR APPLICATION NUMBER: DE 10032529.7
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: DE 10043826.1
: PRIOR FILING DATE: 2000-09-01
: NUMBER OF SEQ ID NOS: 2424
: SEQ ID NO 1670
: LENGTH: 6668
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
: FEATURE:
: NAME/KEY: unsure
: LOCATION: 1936
: OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1670

Query Match      1.5%; Score 62.8; DB 12; Length 6668;
Best Local Similarity 48.1%; Pred. No. 0.00023;

```

	Matches	178;	Conservative	0;	Mismatches	192;	Indels	0;	Gaps	0;
QY	3800	AAAAAGGCGCTCAAAAGACATATATTAGAAAAATGAAGATGAACAGAATAAGTCGGCCAAA								3859
Db	3339	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAACAAAAA								3280
QY	3860	AGGGTAAAGAGGCCGACCACCAAACCTCTTGGTGGAGTACACAAAGAGAGCCAA								3919
Db	3279	AAAAAAAAAACAAAAA								3220
QY	3920	CATATGAACACTTTATAAAAAGAAAGCAAAAAAATCTGGACCTCCAGCACAGAGAGG								3979
Db	3219	AAAAAAAAAAAAACAAAAA								3160
QY	3980	AGGAAGAAAGAAAGAAAGCAAAAGTGAATACGGAACAGAGAGTCCAAAAAGCAACAGACC								4039
Db	3159	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA								3100
QY	4040	GAGTGTCAAGGAGCAGCAGAGAGAGAGCATCTCCGANTCTTAGTGCATTTGAATCCA								4099
Db	3099	CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA								3040
QY	4100	CACAGTCCACACAGAAAGAGAGAGAACCATCCAAAAAGCCATCACATCACAC								4159
Db	3039	AAAAAAAAAAAAAAAAAAAAAAAAACAAAAACAAAAA								2980
QY	4160	CAAAAAAAAA 4169								
Db	2979	AAAAAAAAA 2970								

```

RESULT 13
US-10-001-843-45
; Sequence 45, Application US/10001843
; Publication NO. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Heve
; APPLICANT: Cafferekey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0267
; CURRENT APPLICATION NUMBER: US/10/001.843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 45
; LENGTH: 1267
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (358)..(358)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc.feature
; LOCATION: (478)..(478)
; OTHER INFORMATION: a, c, g or t
US-10-001-843-45

```

Query	Match	Similarity	1.5%:	Score	61.2:	DB	13:	Length	1267:
Best	Local	Similarity	45.9%:	Pred.	No.	0.0002:			
Matches	210:	Conservative	0:	Mismatches	248:	Indels	0:	Gaps	0:
QY	3774	ACAGGACGAAACCTTAAGGCGAGTCAAGCGGCAAAAGAGGCCATACGCGCTTACA							3773
Db	744	AAAAAAAAAAGAGGAAAAACCAACACAAAAAACAACAAAAAAGACACAAAAAA							803
QY	3774	ATCTGATGACAGCAGTGCGCTTGAGGAAAAGAGGCTCAAGACAGATATATTAGAAAAATGA							3833

OY	3197	ACACTGTCGTGAATGTTGCCATGAAATFCAICATGTCAAAGATCTACATACAGTTTGG	3256
Dd	979	AAAAAAAAAATAATTAATTTTTAAAAAAAATAAAAAATAATATTAACAACAAAAA	920
OY	3257	AATCTCCTAAAGACCCTGCTACTACAGCTCGTTCTCTACACACCTGA--CAAGAATT	3314
Dd	919	CCGAAATTAATTAACCTCACACTATATATCTCAATACTTTAAAAAACCTAAATTAATAATTC	860
OY	3315	CAGTAACACAAAAATTAATTCCTCTCGAATGAAATCATTTTTCATCTCGAAAAAC	3374
Dd	859	ACTTAAACGCAAAATTTCAAAACCACTAAACAACAATACTATACCCATTTCACAAA	800
OY	3375	TAAACACCAATGTTCTTAGAGCTGTTAACAGCACTTTCAT--CAGCAGGCAAGCAA	3432
Dd	799	AAAAAAACAAATACTAACCAAAACCTAATTAACACACACACTATATATCCAACTATCAAAA	740
OY	3433	TCTCAGACCAATCATTCAGCAATGAAA--CTGTAGCAATGCAGCAGCAAGCTCAATC	3499
Dd	739	ACTAAACACAAAAAATCTACCTACGCAAAAAACAATTAATTACATTAACCCGAATTCGTACC	680
OY	3491	CAAGCTCTCCCTGGAAATTAAGGGAGGCTTGATAGTTTCGAATGATGCACAGTGAAC	3550
Dd	679	TACACTCCAACTTAACACAAAAAAAACCCCTACTCAAAAAAAAAAATATATTTAATTA	620
OY	3551	ATGAAGATTACACATGTCCTTACCTTTGCCGGGAAAAAAGTGACAAAGAGACGACT	3610

[illegible]

QY 284

Db 98 ATGTTGCGTACTGTAGCCCTGCTGCCATATTTCAGATTTATGTCCTCGAAG 39
QY 284 CTCCTTACACATCCCTGATTAACCTAAG 313
|||||

Db 38 CTCCTTACATCCCTGATAACTAAGG 9

RESULT 2
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 2.4%; Score 99.8; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 2.6e-16;
Matches 17; Conservative 256; Mismatches 118; Indels 0; Gaps 0;

Db 3700 TTGACTAGTGTGTCAGAGACAGAACTAAGCAGCAGAGGAAAGAGGC 3759
1450 TAGAAGATTGTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391
3760 CATACGCTTCAGAACTGATGACAGCAGTGCGCTGAGGAAAGAGCGCTCAAGAGAT 3819
1390 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1331
3820 AATATGAAATGAAGATGACAGATAGTCCGCCAAAAAAGGTTAAAGAGCCGACCA 3879
1330 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1271
3880 CCAAAAGCTCTGTGGAGGTACACCAAGAGCCCAAGCAACTCTAAGAAA 3939
1270 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1211
3940 GGAAGCAAAAAAATCTGACCTCAGCAGCAGAGAGAGAGAAAGAAAGACAA 3999

Db 1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151
4000 AGTGAATATACGACAGAGAGTCCAAAGCAAGCAGCGAGTGTCAAGGAGACAG 4059
1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091
4060 CAGAGCAGATCTCTGAAATCTAGTCAAA 4090
1090 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1060

RESULT 3
US-08-931-999-4
Sequence 4, Application US/08931999
Patent No. 6043219
GENERAL INFORMATION:
APPLICANT: Iandolo, John J.
APPLICANT: Crupper, Scott S.
TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hovey, Williams, Timmons & Collins
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: Missouri
COUNTRY: U.S.A.
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,999
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/710,561
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 25043-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 816/474-9050
TELEFAX: 816/474-9057
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6755 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: UT0007
US-08-931-999-4

Query Match 1.3%; Score 54.6; DB 3; Length 6755;
Best Local Similarity 46.6%; Pred. No. 0.00029;
Matches 174; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

Db 3797 AGGAAAGAGGCTCAAGAGATATATTAGAAATGAAGATCAACAGATAGTCCGCCAA 3856
6204 ACGAAAAACGACAAACAAACAAACAGAGAGAAAAAGAAAAAAGAAAAACGACAGA 6263
3857 AAAAGGTAAAGAGAGCCGACCAACCAACCTCTGTGGAGGTACACCAAAAGAGAGC 3916
6264 GAAACAAAAGCAGGAAACAAACCAAAACCCGCAAGAGAGCAGCAACAAAAGAGAG 6323

[illegible]

```

RESULT 4
US-09-268-992-7/c
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.,
; APPLICANT: Freilmer, N.,
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
; US-09-268-992-7

```

[illegible]

Qy	3993	TTAAAAAGGAGCAAAAAAAATCTGGACCTCCAGACACAGAGGAGGAGAAAGAAAGA	3992
Db	7141	GAAGCAAGGAAGGAAAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGA	7082
Qy	3993	AAGCAAAAGTGGAAATACGGAACAGAAAGTCCAAAAGCAAAAGCAGCCAGTGTGAAGGAG	4052
Db	7081	AAGAAAGCAAGGAATATGAGAAAGAAAGAACGAAAGAGAGAAAGAGAGAGAAAGAA	7022
Qy	4053	AGCACAGCAGAGA	4065
Db	7021	AGAAAGAGAGAGA	7009

```

RESULT 5
US-09-657-474-7/c
Sequence 7, Application US/09657474
Patent No. 6399762
GENERAL INFORMATION:
APPLICANT: Chen, H.
APPLICANT: Prelmer, N.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
FILE REFERENCE: 7853-138
CURRENT APPLICATION NUMBER: US/09/657,474
CURRENT FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: 09/368,992
PRIOR FILING DATE: 1999-03-16
PRIOR APPLICATION NUMBER: 09/236,134
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/106,056
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/088,312
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/078,044
PRIOR FILING DATE: 1998-03-16
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 72604
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified base
LOCATION: all n positions
OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

```

	Query Match	1.2%;	Score 51.4;	DB 4;	Length 72604;
	Best Local Similarity	46.1%;	Pred. No. 0.0076;		
	Matches	172;	Conservative	0;	Mismatches 201; Indels 0; Gaps 0;
OY	3693	GGATGACTTGCTAAGTTGGTACAGGAACAGAAACCTTAAGGCAGTACGCCAGTCCGAA	3752		
Db	7381	GGATTAAGTACGAGCCAGGAGGTTCTTAAGGCTACCTCGAGACATAGTGAAGAACTGGTGGAA	7322		
OY	3753	AAGAGGCCATACGGCTTCAGAAATCGATGAACACACAGTGGCCCTAGAGAAAGAGGCTCAA	3812		
Db	7321	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	7262		
OY	3813	AGAAATATATTAGAAAAATGAAGATGAACAGAAATAGTCCGCCAAAAAAGGTTAAAGAGG	3872		
Db	7261	AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA	7202		
OY	3873	CCGACCAACCAAAACCTCTTGSTGGAGTACCCAAAAAGAGAGCCCAACAATGAAGAACTTC	3932		
Db	7201	GAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA	7142		
OY	3933	TAAAAAAGGAAGCAAAAAAATATCTGCACTCCAGCACAGAGGAGGAGGAAGGAAGGA	3992		
Db	7141	GAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA	7082		
OY	3993	AAGCAAAAGTGGAAATATACGAACGAAGTCCAAAAAGCAAAACGCAACCGAGTGTAAAGAG	4052		


```

1  TITLE OF INVENTION:  AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
2  TITLE OF INVENTION:  DELIVERING POLYNUCLEOTIDE VACCINES
3  NUMBER OF SEQUENCES:  2
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Naval Medical Res. & Dev. Cmd.
6  STREET:  Bldg. 1, T-12 8901 Wisconsin Ave.
7  CITY:  Bethesda
8  STATE:  Maryland
9  COUNTRY:  USA
10 ZIP:  20889-5606
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  Floppy disk
14 COMPUTER:  IBM PC compatible
15 OPERATING SYSTEM:  PC-DOS/MS-DOS
16 SOFTWARE:  PatentIn Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER:  US/08/155,888
19 FILING DATE:
20 CLASSIFICATION:  514
21 ATTORNEY/AGENT INFORMATION:
22 NAME:  Spevack, A. David
23 REGISTRATION NUMBER:  24,743
24 REFERENCE/DOCKET NUMBER:  N.C. 75,851
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE:  (202) 295-6759
27 TELEFAX:  (202) 295-1022
28 INFORMATION FOR SEQ ID NO: 1:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH:  5552 base pairs
31 TYPE:  nucleic acid
32 STRANDEDNESS:  double
33 TOPOLOGY:  circular
34 HYPOTHEICAL:  NO
35 ANTI-SENSE:  NO
36 FEATURE:
37 NAME/KEY:  promoter
38 LOCATION:  1..755
39 IDENTIFICATION METHOD:  experimental
40 OTHER INFORMATION:  /function= "promoter"
41 OTHER INFORMATION:  /evidence= EXPERIMENTAL
42 OTHER INFORMATION:  /label= CMV-IE
43 OTHER INFORMATION:  /note= "this feature acts as a promoter for any
44 OTHER INFORMATION:  downstream DNA sequence."
45 OTHER INFORMATION:  /citation= ([2])
46 FEATURE:
47 NAME/KEY:  CDS
48 LOCATION:  933..2367
49 IDENTIFICATION METHOD:  experimental
50 OTHER INFORMATION:  /codon_start= 933
51 OTHER INFORMATION:  /function= "protein protective against malaria"
52 OTHER INFORMATION:  /product= "protein"
53 OTHER INFORMATION:  /evidence= EXPERIMENTAL
54 OTHER INFORMATION:  /number= 1
55 OTHER INFORMATION:  /label= IL2-CSP
56 OTHER INFORMATION:  /citation= ([1])
57 PUBLICATION INFORMATION:
58 AUTHORS:  Sedegah, Martha
59 AUTHORS:  Hedstrom, Richard C.
60 AUTHORS:  Hoffman, Stephen L.
61 TITLE:  Vaccination with Plasmodium yoelii CS protein
62 TITLE:  Plasmid DNA protects against malaria
63 JOURNAL:  Science
64 PUBLICATION INFORMATION:
65 AUTHORS:  Cullen, Bryan R.
66 TITLE:  TRANS-ACTIVATION OF HUMAN IMMUNODEFICIENCY
67 TITLE:  VIRUS OCCURS VIA A BIMODAL MECHANISM
68 JOURNAL:  CELL
69 VOLUME:  46
70 PAGES:  973-982
71 DATE:  26 SEP-1986
72 RELEVANT RESIDUES IN SEQ ID NO: 1:  FROM 1 TO 4732
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872

```

[illegible]

```

RESULT 12
US-09-434-408-3
: Sequence 3, Application US/09434408
: Patent No. 6440697
: GENERAL INFORMATION:
: APPLICANT: Venezia, Domenick
: APPLICANT: Grossmann, Angelika
: TITLE OF INVENTION: RING FINGER PROTEIN ZAPOP3
: FILE REFERENCE: 98-41
: CURRENT APPLICATION NUMBER: US/09/434,408
: CURRENT FILING DATE: 1999-11-04
: EARLIER APPLICATION NUMBER: US 60/108,258
: EARLIER FILING DATE: 1998-11-12
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 2169
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Degenerate polynucleotide sequence of zapop3
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(2169)
: OTHER INFORMATION: n = A,T,C OR G
: US-09-434-408-3

```

[illegible]

FT	exon	378..464
FT		/number= e
FT	exon	/number= 4
FT		465..562
FT		/tag= f
FT	exon	/number= 5
FT		563..669
FT		/tag= e
FT	exon	/number= 6
FT		690..854
FT		/tag= f
FT	exon	/number= 7
FT		855..911
FT		/tag= g
FT	exon	/number= 8
FT		912..1027
FT		/tag= h
FT	exon	/number= 9
FT		1028..1122
FT		/tag= i
FT	exon	/number= 10
FT		1123..1268
FT		/tag= j
FT	exon	/number= 11
FT		1269..1420
FT		/tag= k
FT	exon	/number= 12
FT		1421..1534
FT		/tag= l
FT	exon	/number= 13
FT		1535..1616
FT		/tag= m
FT	exon	/number= 14
FT		1617..1665
FT		/tag= n
FT	exon	/number= 15
FT		1666..1805
FT		/tag= o
FT	exon	/number= 16
FT		1806..1921
FT		/tag= p
FT	exon	/number= 17
FT		1922..2027
FT		/tag= q
FT	exon	/number= 18
FT		2028..2188
FT		/tag= r
FT	exon	/number= 19
FT		2189..2312
FT		/tag= s
FT	exon	/number= 20
FT		2313..2471
FT		/tag= t
FT	exon	/number= 21
FT		2472..2540
FT		/tag= u
FT	exon	/number= 22
FT		2541..2677
FT		/tag= v
FT	exon	/number= 23
FT		2678..2801
FT		/tag= w
FT	exon	/number= 24
FT		2802..3006
FT		/tag= x
FT	exon	/number= 25
FT		3007..3121
FT		/tag= y
FT	exon	/number= 26
FT		3122..3254
FT		/tag= z
FT	exon	/number= 27
FT		3255..3374

FT		/tag= aa
FT	/number= 28	
FT	3375..3437	
FT	/tag= ab	
FT	/number= 29	
FT	3438..3583	
FT	/tag= ac	
FT	/number= 30	
FT	3584..3689	
FT	/tag= ad	
FT	/number= 31	
FT	3690..4129	
FT	/tag= ae	
FT	/number= 32	
FT	4130..4354	
FT	/tag= af	
FT	/number= 33	
FT	4320..4324	
FT	/tag= ag	
FT	/note= "Destabilizing signal"	
FT	4355..5253	
FT	/tag= ah	
FT	/number= 34	
FT	5228..5233	
FT	/tag= ai	
FT	5249..5253	
FT	/tag= aj	
FT	/note= "Poly adenylation cleavage site"	
XX		
FN	W0200050454-A1.	
XX		
XX	31-AUG-2000.	
PD		
XX		
PE	24-FEB-2000; 2000WO-US04732.	
XX		
PR	24-FEB-1999; 99US-0121461.	
XX		
PA	(TUFFT) TUFFTS COLLEGE.	
XX		
P1	Soto AM, Sonnenschein C, Geck P, Szelel J;	
DR	WPI: 2000-565451/52.	
XX	P-PSDB; AARY94702.	
PT		
PT	New human androgen-induced tumor suppressor cDNA sequence termed	
PT	'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient	
PT	diagnosis and treatment of prostate cancer -	
XX		
PS	Claim 1; Fig 1; 152pp; English.	
XX		
CC	This invention relates to a human androgen-induced tumour suppressor cDNA	
CC	sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located	
CC	on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell	
CC	proliferation and use as a marker for the efficient diagnosis and	
CC	treatment of prostate cancer. The invention includes AS3 cDNA and protein	
CC	sequences, a vector comprising the cDNA sequence, a host cell transfected	
CC	with the expression vector, and a method for producing an AS3 polypeptide	
CC	comprising culturing the transfected cells. AS3 has cytostatic activity,	
CC	and acts to suppress cell proliferation. The AS3 gene is useful as a	
CC	marker for the efficient diagnosis and treatment of prostate cancer. The	
CC	AS3 nucleic acid molecule can be used as a source of antisense agents for	
CC	sequence specific modulation of gene expression. The AS3 protein may be	
CC	used in the treatment of disorders caused by aberrant modification or	
CC	mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene	
CC	or aberrant post-translational modification of the AS3 protein. This	
CC	sequence represents the human AS3 cDNA sequence.	
XX		
SQ	Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;	
Query Match	100.0%; Score 4173; DB 21; Length 5271;	
Best Local Similarity	100.0%; Pred. NO. 0;	
Matches 4173; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	

Db	1146	GACCTCAGGAAGCTATTAGACATGATGTTATTGTGTCAATATTAACACTCTCTAAAG	1205
OY	1141	GATATTTCTTGCGTCATCATGATCACTACTTAATTTTGTGAGAGAGAACAATTAGCAAA	1200
Db	1206	GATTTTCTTCTGGTCATATGATCACTTAATTTTGTGAGAGAGAAACATTAGACAAA	1265
OY	1201	CGATGGAGAGTACGCCAAAAGCCATGATGGAGCTTGCCCAATTTATTAAGAAATATGCT	1260
Db	1266	CGATGGAGAGTACGCCAAAAGCCATGATGGAGCTTGCCCAATTTATTAAGAAATATGCT	1325
OY	1261	TTTACATGCACAGCTGGAAAAAGATGCTGCAAAACAGATAGCATGGATCTAAAGCAAAATTG	1320
Db	1326	TTTACATGCACAGCTGGAAAAAGATGCTGCAAAACAGATAGCATGGATCTAAAGCAAAATTG	1385
OY	1321	CTACATATATATTTATTCAAAATPAGTATTTAGATGCATCACTTGTTAACAGATCTTTGCT	1380
Db	1386	CTACATATATATTTATTCAAAATPAGTATTTAGATGCATCACTTGTTAACAGATCTTTGCT	1445
OY	1446	CAATACATATGTTCTCTCAATTTAGAAAACACAGAAAGATGTAATGCTTATATTAATCTTG	1440
Db	1441	TATCCACACTGGATTTAAATGCTGTGAAAGCATTTGAATGAATGTGGAAATGTCAAAAT	1500
OY	1501	CTGCTCGACATCAAGTAAAGATTTGCTTGATTTAGCAACCCAAAACAGATGTC	1560
Db	1566	CTGCTCGACATCAAGTAAAGATTTGCTTGATTTAGCAACCCAAAACAGATGTC	1625
OY	1561	AGTGTCAAGGCCATATTTTCAAAAAGTATGATGTTATTTACAGAAATTTACCTGATCTGT	1620
Db	1626	AGTGTCAAGGCCATATTTTCAAAAAGTATGATGTTATTTACAGAAATTTACCTGATCTGT	1685
OY	1621	AAGCTCAGATTTTATGAAAGAAATTTACACAGGTGTAGAGATGATAGAAATATAGA	1680
Db	1686	AAGCTCAGATTTTATGAAAGAAATTTACACAGGTGTAGAGATGATAGAAATATAGA	1745
OY	1681	AAGCAGTTAGAGTACTGTTTATGCCAACAATGCTCTGCAAGAGCTGAAGTGTGTG	1740
Db	1746	AAGCAGTTAGAGTACTGTTTATGCCAACAATGCTCTGCAAGAGCTGAAGTGTGTG	1805
OY	1741	CGTGAATTAATCTAAGAAAGTTGGGCAACCCCAACAGCCTTACAAATCTTTCTCGAATG	1800
Db	1806	CGTGAATTAATCTAAGAAAGTTGGGCAACCCCAACAGCCTTACAAATCTTTCTCGAATG	1865
OY	1801	ATCAAGTTTCTCTTGGAGAGATAGCACCTGTGCATATAGATACGAATCTATCAGTCT	1860
Db	1866	ATCAAGTTTCTCTTGGAGAGATAGCACCTGTGCATATAGATACGAATCTATCAGTCT	1925
OY	1861	CTTATTTAAACAAGTGAACAATCAATATAGTGAACAAGCAGATGATTAACATAGAGGTCT	1920
Db	1926	CTTATTTAAACAAGTGAACAATCAATATAGTGAACAAGCAGATGATTAACATAGAGGTCT	1985
OY	1921	CCAACGTATCAAGGCATCAGCAGAGCTTGTAACCTCTTAAGCTACTCATTTACACAT	1980
Db	1986	CCAACGTATCAAGGCATCAGCAGAGCTTGTAACCTCTTAAGCTACTCATTTACACAT	2045
OY	1981	CCCATCTCATTTGATTTCTGCTGAACAATTTGAATTCATCTACGTGCTGTGAAAAATGAT	2040
Db	2046	CCCATCTCATTTGATTTCTGCTGAACAATTTGAATTCATCTACGTGCTGTGAAAAATGAT	2105
OY	2041	GATGAAAAAATAGCAGAGCTGCACATCAAAATTTTAAAAAACACAGAAAGCAAAATTTGAA	2100
Db	2106	GATGAAAAAATAGCAGAGCTGCACATCAAAATTTTAAAAAACACAGAAAGCAAAATTTGAA	2165
OY	2101	GAGGATTTTTCACACATCAGATCAGCTTGCTCTCGTTTTCATATACAAATCTAAAAAA	2160
Db	2166	GAGGATTTTTCACACATCAGATCAGCTTGCTCTCGTTTTCATATACAAATCTAAAAAA	2225
OY	2161	GGACCCCCCGTCAAGCCAAATATGCCATTCATTTGATTCGATATATTTTCTAGTAAA	2220

Db	2226	GGACCCCCCGCTCAAGCCAAATATGCCATTGATGTATTCATGCGATGGATATTTTCTAGTAA	2285
OY	2221	GAGACCCAGTTTGCACAGATATTTGAGCTCTGCTATGAAGCCTTAGATCCAGCAACCTG	2280
Db	2286	GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCTATGAAGCCTTAGATCCAGCAACCTG	2345
OY	2281	GAACATCTCAATAACACATGGTGTACATATGGTATATATGCTCTCCATTGCAACCGATCA	2340
Db	2346	GAACATCTCAATAACACATGGTGTACATATGGTGTATATGCTCTCCATTGCAACCGATCA	2405
OY	2341	TTTGCTGCTCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATAA	2400
Db	2406	TTTGCTGCTCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATAA	2465
OY	2401	GATGCGCTCCAGGGAAAAAGACAACCTAACTTTGGTTCAGATGGAAGATATCTCT	2460
Db	2466	GATGCGCTCCAGGGAAAAAGACAACCTAACTTTGGGTTCAGATGGAAGATATCTCT	2525
OY	2461	GAGACAAATGGTCAAAATTCAGGCTATTTAAATGATGGTGTGATGGTCACTTGGATGAA	2520
Db	2526	GAGACAAATGGTCAAAATTCAGGCTATTTAAATGATGGTGTGATGGTCACTTGGATGAA	2585
OY	2521	AATATCAACAGATAATCAGAACTTCTACCTTAAAGATTGCTAAACAAACATATTCATAGT	2580
Db	2586	AATATCAACAGATAATCAGAACTTCTACCTTAAAGATTGCTAAACAAACATATTCATAGT	2645
OY	2581	GATGAGACTTGCACAGAACAGGGAAAAATTAGTAACCCAGATATGACGCTTGAGACTT	2640
Db	2646	GATGAGACTTGCACAGAACAGGGAAAAATTAGTAACCCAGATATGACGCTTGAGACTT	2705
OY	2641	GCTGCTGGAGATGCTATTTGTGAACTGGCACAAACCCGTTCACATGAAATATACACA	2700
Db	2706	GCTGCTGGAGATGCTATTTGTGAACTGGCACAAACCCGTTCACATGAAATATACACA	2765
OY	2701	TTAGAACAAATATCAGCTATGTGCATTAGCTATCAACGATGATGCTATCAAGTAAGCAA	2760
Db	2766	TTAGAACAAATATCAGCTATGTGCATTAGCTATCAACGATGATGCTATCAAGTAAGCAA	2825
OY	2761	GTGTTTGGCCCAAAACCTCACAAAGGCGTTTCCGCTTACGGCTTCCACTTGAATATG	2820
Db	2826	GTGTTTGGCCCAAAACCTCACAAAGGCGTTTCCGCTTACGGCTTCCACTTGAATATG	2885
OY	2821	GCAATCTGTGGCCCTTGTGCAAAGATCCGTAAAGGAGAAACAGTCATGCAAGGCAA	2880
Db	2886	GCAATCTGTGGCCCTTGTGTGCAAAGATCTTGAAGGAGAAAGCTCATGCTAAGGCA	2945
OY	2881	TGTTTGGTGA AAAATATTAATGTAAAGGGGAGATCTGAAGCAGCATGCACTGTTAGT	2940
Db	2946	TGTTTGGTGA AAAATATTAATGTAAAGGGGAGATCTGAAGCAGCATGCACTGTTAGT	3005
OY	2941	GAAAAATATATGTCTCTTCTACACAGATATGTTTCCATATACAAATTCACCTTTGGCA	3000
Db	3006	GAAAAATATATGTCTCTTCTACACAGATATGTTTCCATATACAAATTCACCTTTGGCA	3065
OY	3001	CATGACCCAGATTTATGTCAAAGTACAGATATTTGAACAACCTAAAGATTTAAAGATGT	3060
Db	3066	CATGACCCAGATTTATGTCAAAGTACAGATATTTGAACAACCTAAAGATTTAAAGATGT	3125
OY	3061	CTTTGGTTTGTCTGCGAATATTAATGAGCTAAAAATGAATTAACAGTCAACGCTTTTATC	3120
Db	3126	CTTTGGTTTGTCTGCGAATATTAATGAGCTAAAAATGAATTAACAGTCAACGCTTTTATC	3185
OY	3121	AGAAAGATGTAGAAAAATTTTAAACAAACAAAGATGCCAAGGACCAAGATGATGCAAA	3180
Db	3186	AGAAAGATGTAGAAAAATTTTAAACAAACAAAGATGCCAAGGACCAAGATGATGCAAA	3245
OY	3181	ATGAATGAAAAACGTACACTGTGTGTATGTTGCGATGAATATATCATGCTCAAAAGT	3240
Db	3246	ATGAATGAAAAACGTACACTGTGTGTATGTTGCGATGAATATATCATGCTCAAAAGT	3305
OY	3241	ACTACATACAGTTGGATCTCTTAACAAACCCGGACTACACAGCTCGTTTCTCACTCAA	3300
Db	3306	ACTACATACAGTTGGATCTCTTAACAAACCCGGACTACACAGCTCGTTTCTCACTCAA	3365

QY	3301	CCTGACAAAGATTTTCAGTAACACCACCAAAATTAATCTGGCTCTCTGAAATGAATCAATTTTC	3360
Db	3366	CCTGACAAAGATTTTCAGTAACACCACCAAAATTAATCTGGCTCTCTGAAATGAATCAATTTTC	3425
QY	3361	ACTCTGGAAAACCTAAACCAACCAATGTTTGTAGAGCTGTGTAACAAGCCACTTTCATCA	3420
Db	3426	ACTCTGGAAAACCTAAACCAACCAATGTTTGTAGAGCTGTGTAACAAGCCACTTTCATCA	3485
QY	3421	GCAGGCAAGCAATCTCAGACCAAAATCATCAGCAATGGAAAACCTGTAGCAATGCAGAGC	3480
Db	3486	GCAGGCAAGCAATCTCAGACCAAAATCATCAGCAATGGAAAACCTGTAGCAATGCAGAGC	3545
QY	3481	AGCTCAATTCCAAGCTCTCTTGAAAGATAAAGGGAGGCTTGATTAATTCTGAAATGGAT	3540
Db	3546	AGCTCAATTCCAAGCTCTCTTGAAAGATAAAGGGAGGCTTGATTAATTCTGAAATGGAT	3605
QY	3541	CACAGTGAATTAAGATTAACACATGCTTCACCTTTGGCCGGGGAAAAAAGAGCAAG	3600
Db	3606	CACAGTGAATTAAGATTAACACATGCTTCACCTTTGGCCGGGGAAAAAAGAGCAAG	3665
QY	3601	AGAGACGACTCTGATCTTTGAAGTCTGAAATGGAGAAAGCCTTAGAGCAGAGAAAAAACG	3660
Db	3666	AGAGACGACTCTGATCTTTGAAGTCTGAAATGGAGAAAGCCTTAGAGCAGAGAAAAAACG	3725
QY	3661	CCCGTCACAGAACAGAGAGAAATTAAGTATGATGACTTGACTTAAGTTGCTACAGAA	3720
Db	3726	CCCGTCACAGAACAGAGAGAAATTAAGTATGATGACTTGACTTAAGTTGCTACAGAA	3785
QY	3721	CAGAAACCTTAAGGAGTCAAGCAAGTGGAAAAAGAGCCATTAAGGCTTCAGATCTGAT	3780
Db	3786	CAGAAACCTTAAGGAGTCAAGCAAGTGGAAAAAGAGCCATTAAGGCTTCAGATCTGAT	3845
QY	3781	GAACGACAGTGGCTGAGAGAAAAGAGGCTCAAAAGATATATTAGAAAATGAAGATGAA	3840
Db	3846	GAACGACAGTGGCTGAGAGAAAAGAGGCTCAAAAGATATATTAGAAAATGAAGATGAA	3905
QY	3841	CAGATTAATCCGCCAAAAAAGGTTAAAGAGCCGACCCACCAAAACCTTGTGGGAGGT	3900
Db	3906	CAGATTAATCCGCCAAAAAAGGTTAAAGAGCCGACCCACCAAAACCTTGTGGGAGGT	3965
QY	3901	ACACCAAAAGAAAGAGCCACCAATGAAACTTCTAAAAAAGGAGCAAAAAAATTCGTGA	3960
Db	3966	ACACCAAAAGAAAGAGCCACCAATGAAACTTCTAAAAAAGGAGCAAAAAAATTCGTGA	4025
QY	3961	CCTCCAGCAGCAGAGGAGGAGAAAGAAAGAAAGACAAAGTGGAAATACGGAACAGAG	4020
Db	4026	CCTCCAGCAGCAGAGGAGGAGAAAGAAAGAAAGACAAAGTGGAAATACGGAACAGAG	4085
QY	4021	TCCAAAAGCAAAACAGCACCGAGTGTCAAGGAGACACAGCAGAGCAAGATCTCTGAA	4080
Db	4086	TCCAAAAGCAAAACAGCACCGAGTGTCAAGGAGACACAGCAGAGCAAGATCTCTGAA	4145
QY	4081	TCTAGTGAATTGAATTCACACAGTCCACACACAGAAAGAGACGAGAAAGCATCAAAA	4140
Db	4146	TCTAGTGAATTGAATTCACACAGTCCACACACAGAAAGAGACGAGAAAGCATCAAAA	4205
QY	4141	ACGCCATCAGCATCACAACCAAAAAAATATGTG 4173	
Db	4206	ACGCCATCAGCATCACAACCAAAAAAATATGTG 4238	
RESULT 2			
ID	AAA28052	standard; cDNA; 5355 BP.	
AC	AAA28052;		
XX			
DT	01-DEC-2000	(first entry)	
XX			
DE	Human androgen shutoff gene 3 (AS3)	cDNA sequence SEQ ID #4.	
XX			
KW	Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;		

KW	chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
XW	diagnosis; treatment: cytostatic; human; ss.
XX	Homo sapiens.
FT	Key
FT	Location/Qualifiers
FT	exon
FT	/tag= a
FT	/number= 1
FT	131..257
FT	/tag= b
FT	/number= 2
FT	150..4325
FT	CDS
FT	/tag= c
FT	/product= "AS3"
FT	/note= "Androgen shutoff gene 3 protein"
FT	258..461
FT	exon
FT	/tag= d
FT	/number= 3
FT	462..548
FT	/tag= e
FT	/number= 4
FT	549..646
FT	/tag= f
FT	/number= 5
FT	647..773
FT	/tag= e
FT	/number= 6
FT	774..938
FT	/tag= f
FT	/number= 7
FT	939..995
FT	/tag= g
FT	/number= 8
FT	996..1111
FT	/tag= h
FT	/number= 9
FT	1112..1206
FT	/tag= i
FT	/number= 10
FT	1207..1352
FT	/tag= j
FT	/number= 11
FT	1353..1504
FT	/tag= k
FT	/number= 12
FT	1505..1618
FT	/tag= l
FT	/number= 13
FT	1619..1700
FT	/tag= m
FT	/number= 14
FT	1701..1749
FT	/tag= n
FT	/number= 15
FT	1750..1889
FT	/tag= o
FT	/number= 16
FT	1890..2005
FT	/tag= p
FT	/number= 17
FT	2006..2111
FT	/tag= q
FT	/number= 18
FT	2112..2272
FT	/tag= r
FT	/number= 19
FT	2273..2396
FT	/tag= s
FT	/number= 20
FT	2397..2555
FT	/tag= t
FT	/number= 21

FT	exon	2556..2624
FT		/tag= u
FT		/number= 22
FT	exon	2625..2761
FT		/tag= v
FT		/number= 23
FT	exon	2672..2885
FT		/tag= w
FT		/number= 24
FT	exon	2886..3090
FT		/tag= x
FT		/number= 25
FT	exon	3091..3205
FT		/tag= y
FT		/number= 26
FT	exon	3206..3338
FT		/tag= z
FT		/number= 27
FT	exon	3339..3458
FT		/tag= aa
FT		/number= 28
FT	exon	3459..3521
FT		/tag= ab
FT		/number= 29
FT	exon	3522..3667
FT		/tag= ac
FT		/number= 30
FT	exon	3668..3773
FT		/tag= ad
FT		/number= 31
FT	exon	3774..4213
FT		/tag= ae
FT		/number= 32
FT	exon	4214..4438
FT		/tag= af
FT		/number= 33
FT	misc_signal	4314..4408
FT		/tag= ag
FT		/note= "Destabilizing signal"
FT		4439..5337
FT	exon	/tag= ah
FT		/number= 34
FT	polyA_signal	5312..5317
FT		/tag= ai
FT		5333..5337
FT	misc_feature	/tag= aj
FT		/note= "Poly adenylation cleavage site"
PN		WO200050454-A1.
PN		
PD		31-AUG-2000.
XX		
XX		
PF		24-FEB-2000; 2000WO-US04732.
XX		
PR		24-FEB-1999; 99US-0121461.
XX		
PA		(TUFT) TUFTS COLLEGE.
XX		
PI		Soto AM, Sonnenschein C, Geck P, Szelei J;
XX		
DR		WPI; 2000-565451/52.
DR		P-PSDB; MAY94702.
XX		
PT		New human androgen-induced tumor suppressor cDNA sequence termed
PT		'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT		diagnosis and treatment of prostate cancer -
XX		
PS		Example 4; Fig 6; 152pp; English.
XX		
CC		This invention relates to a human androgen-induced tumour suppressor
CC		cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
CC		located on chromosome 13 at position 13q12-13q. AS3 has a role in
CC		inhibiting cell proliferation and use as a marker for the efficient

CC diagnosis and treatment of prostate cancer. The invention includes AS3
CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host
CC cell transfected with the expression vector, and a method for producing
CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC cytoskeletal activity, and acts to suppress cell proliferation. The AS3
CC gene is useful as a marker for the efficient diagnosis and treatment of
CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC antisense agents for sequence specific modulation of gene expression. The
CC AS3 protein may be used in the treatment of disorders caused by aberrant
CC modification or mutation of a gene encoding an AS3 protein, misregulation
CC of the AS3 gene or aberrant post-translational modification of the AS3
CC protein. This sequence represents the human AS3 cDNA sequence with an
CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC compared with the claimed AS3 cDNA sequence AAA28051.

XX Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 4173; DB 21; Length 5355;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGCTCATTTCAAGACTAGACCAATGATGGAATAATTACATATCCGCTGGGCTCAAG 60
DB |||||||
OY 150 ATGGCTCATTTCAAGACTAGACCAATGATGGAATAATTACATATCCGCTGGGCTCAAG 209
DB |||||||
OY 61 GAAATATCAGATTAATAATCTTAAAGAGAGATGGTGAGACGATTAAAGATGTTTGAAA 120
DB |||||||
OY 210 GAAATATCAGATTAATAATCTTAAAGAGAGATGGTGAGACGATTAAAGATGTTTGAAA 269
DB |||||||
OY 121 ACTTTTATGATATGACACGAGACTCTGTAAGAAAAGAGCTTTATTTAACTTACT 180
DB |||||||
OY 270 ACTTTTATGATATGACACGAGACTCTGTAAGAAAAGAGCTTTATTTAACTTACT 329
DB |||||||
OY 181 TTACATCTTGCTTATGATATTTTCTCAGACATCTGTGTAAGATGTTGCTTACTGTGA 240
DB |||||||
OY 330 TTACATCTTGCTTATGATATTTTCTCAGACATCTGTGTAAGATGTTGCTTACTGTGA 389
DB |||||||
OY 241 GCGTGTGCTTGTCTGATATTTTTCAGATTTATGCTCTGAGCTCTTTACACATCCCT 300
DB |||||||
OY 390 GCGTGTGCTTGTCTGATATTTTTCAGATTTATGCTCTGAGCTCTTTACACATCCCT 449
DB |||||||
OY 301 GATAAATCAAGGATATTTTATGTTTATACAGACAGTGAAGGGCTAGAGATACA 360
DB |||||||
OY 450 GATAAATCAAGGATATTTTATGTTTATACAGACAGTGAAGGGCTAGAGATACA 509
DB |||||||
OY 361 AAGAGCCCAATTCATAGTATTTTATTTTACTTGAACATTTGCTTGGGTCAAGTCA 420
DB |||||||
OY 510 AAGAGCCCAATTCATAGTATTTTATTTACTTGAACATTTGCTTGGGTCAAGTCA 569
DB |||||||
OY 421 TATAACATATGCTTTGAGTATGAGATAGCAATGAAATTTTCCACCAGCTATACAGAAC 480
DB |||||||
OY 570 TATAACATATGCTTTGAGTATGAGATAGCAATGAAATTTTCCACCAGCTATACAGAAC 629
DB |||||||
OY 481 TTATTTTTCAGTATTAACAAATGCGCCACAATGCAATATGACATAGTACACTT 540
DB |||||||
OY 630 TTATTTTTCAGTATTAACAAATGCGCCACAATGCAATATGACATAGTACACTT 689
DB |||||||
OY 541 ATGAGCTCTATTTTGTGAAGGTGATPACAGTGTCTCAGAGCTTTTGGATACGGTTTA 600
DB |||||||
OY 690 ATGAGCTCTATTTTGTGAAGGTGATPACAGTGTCTCAGAGCTTTTGGATACGGTTTA 749
DB |||||||
OY 601 GTAATCTGTACTGCTCATTAAGAAATTTAAACAAGCAATATGATTTGGCAAAAGCT 660
DB |||||||
OY 750 GTAATCTGTACTGCTCATTAAGAAATTTAAACAAGCAATATGATTTGGCAAAAGCT 809
DB |||||||
OY 661 TTACTGAAGAGACAGCTCAAGCTATTTAGCCATATATACCACTTTTAAATACAGT 720
DB |||||||
OY 810 TTACTGAAGAGACAGCTCAAGCTATTTAGCCATATATATACCACTTTTAAATACAGT 869
DB |||||||
OY 721 CTGATGCTTGGGAAACATCTATCAGAGATTTGTCAGAGCAATGCTTGAATATTTG 780
DB |||||||
OY 870 CTGATGCTTGGGAAACATCTATCAGAGATTTGTCAGAGCAATGCTTGAATATTTG 929
DB |||||||

OY 781 GAGCTCTACATATATGATAGTCAATTTGCTCTCTCTGTGTTTACCACCAGCTTGAATTTAA 840
DB |||||||
OY 930 GAGCTCTACATATATGATAGTCAATTTGCTCTCTCTGTGTTTACCACCAGCTTGAATTTAA 989
DB |||||||
OY 841 TTAAAGGCAATGATATGAGAGCGCTTCAAGTTGTTTAACTACTGCGAAAATGTTT 900
DB |||||||
OY 990 TTAAAGGCAATGATATGAGAGCGCTTCAAGTTGTTTAACTACTGCGAAAATGTTT 1049
DB |||||||
OY 901 GGGGCAAGGATTCAGATATGGCTTCTCAAAACAGCCACTTTGGCAGTCTTACTGGG 960
DB |||||||
OY 1050 GGGGCAAGGATTCAGATATGGCTTCTCAAAACAGCCACTTTGGCAGTCTTACTGGG 1109
DB |||||||
OY 961 AGTTTATGATATTCATGTAACCAATCCGCTGGAATGTGTGAATTTGCTAGCCATTGT 1020
DB |||||||
OY 1110 AGTTTATGATATTCATGTAACCAATCCGCTGGAATGTGTGAATTTGCTAGCCATTGT 1169
DB |||||||
OY 1021 CTCATGAACCATCCGTGATTTAGCAAAAGACTTAAACAGTATCTTAAAGTGAAGTCAAT 1080
DB |||||||
OY 1170 CTCATGAACCATCCGTGATTTAGCAAAAGACTTAAACAGTATCTTAAAGTGAAGTCAAT 1229
DB |||||||
OY 1081 GACCCGTAGAGAGCTATTTAGCAATGATGTTATGTGTCAATATGATTAACGCTGCTAATAAG 1140
DB |||||||
OY 1230 GACCCGTAGAGAGCTATTTAGCAATGATGTTATGTGTCAATATGATTAACGCTGCTAATAAG 1289
DB |||||||
OY 1141 GATATTTCTTGTGCTCAATGATCACTTACTTATTTTGTGAGAGAGAAACATTAAGCAAA 1200
DB |||||||
OY 1290 GATATTTCTTGTGCTCAATGATCACTTACTTATTTTGTGAGAGAGAAACATTAAGCAAA 1349
DB |||||||
OY 1201 CGATGAGAGATAGCAAAAGAGCCATGATGGAGCTTCCCAAAATTTTAAAGAAATATGCT 1260
DB |||||||
OY 1350 CGATGAGAGATAGCAAAAGAGCCATGATGGAGCTTCCCAAAATTTTAAAGAAATATGCT 1409
DB |||||||
OY 1261 TTACAGTCAGAGCTGGAAGAAAGATGTCGCAAAACAGATAGCATGATCAAAAGCAAAATG 1320
DB |||||||
OY 1410 TTACAGTCAGAGCTGGAAGAAAGATGTCGCAAAACAGATAGCATGATCAAAAGCAAAATG 1469
DB |||||||
OY 1321 CTACATATATATTTCAAAATATGATATGATGACTACTGTTGTAAGAGATTTTGTCT 1380
DB |||||||
OY 1470 CTACATATATATTTCAAAATATGATATGATGACTACTGTTGTAAGAGATTTTGTCT 1529
DB |||||||
OY 1381 CAATACATATGTTCTCTCAAAATTTGAAGATCAAGACGATGAAATGCTATATTTACTTGT 1440
DB |||||||
OY 1330 CAATACATATGTTCTCTCAAAATTTGAAGATCAAGACGATGAAATGCTATATTTACTTGT 1589
DB |||||||
OY 1441 TATGCCACACTGGAATTTTAAATGCTGTGAAGCAATTTGAATGTTGAAATGTCMAAAT 1500
DB |||||||
OY 1590 TATGCCACACTGGAATTTTAAATGCTGTGAAGCAATTTGAATGTTGAAATGTCMAAAT 1649
DB |||||||
OY 1501 CTGCTCGACATCAAGTAAAGATTTGCTGTGACTTGTATTAAGCAACCCAAACAGATGCC 1560
DB |||||||
OY 1650 CTGCTCGACATCAAGTAAAGATTTGCTGTGACTTGTATTAAGCAACCCAAACAGATGCC 1709
DB |||||||
OY 1561 AGTGCAAGGCCATATTTTCAAAAGTGTGTTTCAAGAAATTTTACCGTATCCGTG 1620
DB |||||||
OY 1710 AGTGCAAGGCCATATTTTCAAAAGTGTGTTTCAAGAAATTTTACCGTATCCGTG 1769
DB |||||||
OY 1621 AAGGCTCAGATTTTCATGAAGAAATTCACACAGGCTGTAGAAGATGATGAGAAATTAAGA 1680
DB |||||||
OY 1770 AAGGCTCAGATTTTCATGAAGAAATTCACACAGGCTGTAGAAGATGATGAGAAATTAAGA 1829
DB |||||||
OY 1681 AAGCAGTTAGAGTACTTGTATGTCCAACATGCTCTGTCGAAGCAGGCTGAAGGTTGTGTG 1740
DB |||||||
OY 1830 AAGCAGTTAGAGTACTTGTATGTCCAACATGCTCTGTCGAAGCAGGCTGAAGGTTGTGTG 1889
DB |||||||
OY 1741 CGTGAATTAATCTAAGAGTTGGGACCCCAAGAGCTACAAATTCCTTCCGTGAAGATG 1800
DB |||||||
OY 1890 CGTGAATTAATCTAAGAGTTGGGACCCCAAGAGCTACAAATTCCTTCCGTGAAGATG 1949
DB |||||||
OY 1801 ATCAAGTTTCTCTTGGAGAGATAGCAACCTGTGCAATATGATACCAATCTATCACTGCT 1860
DB |||||||
OY 1950 ATCAAGTTTCTCTTGGAGAGATAGCAACCTGTGCAATATGATACCAATCTATCACTGCT 2009
DB |||||||
OY 1861 CTATTTAAACAAGTGAACAAATCAATAGATGAAGACAGCATGATGAAGATGAGGTTGT 1920
DB |||||||

Db 2010 CTTATTAAACAAAGTGAACAAATCAATAGATGGAACACAGATGATGAAGATGAGGGTGT 2069
QY 1921 CCAAGTATCAAGCATCAGAGAGGCTTGAAGCTTTAAGGACTCTCATTTACAT 1980
Db 2070 CCAAGTATCAAGCATCAGAGAGGCTTGAAGCTTTAAGGACTCTCATTTACAT 2129
QY 1981 CCCATCTCATTTCTTCTGTAACATTTGAATCATTTACTGGCTGTCTGAANAATGAT 2040
Db 2130 CCCATCTCATTTCTTCTGTAACATTTGAATCATTTACTGGCTGTCTGAANAATGAT 2189
QY 2041 GATGAAAAAGTAGCAGAGCTGCATTCAAATTTTCAAAAACACAGAGCAAAATTTGAA 2100
Db 2190 GATGAAAAAGTAGCAGAGCTGCATTCAAATTTTCAAAAACACAGAGCAAAATTTGAA 2249
QY 2101 GAGGATTTTCCACACATCAGATCAGCTTGTCTCCGTGTTTACTATCAATCAATCTAAAAAA 2160
Db 2250 GAGGATTTTCCACACATCAGATCAGCTTGTCTCCGTGTTTACTATCAATCTAAAAAA 2309
QY 2161 GGAACCCCGCGTCAGGCAAAATATGCCATTCAATTTATCCATGCGATATTTTCTAGTAA 2220
Db 2310 GGAACCCCGCGTCAGGCAAAATATGCCATTCAATTTATCCATGCGATATTTTCTAGTAA 2369
QY 2221 GAGACCCAGTTTGCACAGATTTTGAAGCTCTGCATTAAGGCTTATCCAGCAACCTG 2280
Db 2370 GAGACCCAGTTTGCACAGATTTTGAAGCTCTGCATTAAGGCTTATCCAGCAACCTG 2429
QY 2281 GAACATCTCATTAACACATGTTTACTATTGGTCATATTGCTCCCTTGCACCGTATCA 2340
Db 2430 GAACATCTCATTAACACATGTTTACTATTGGTCATATTGCTCCCTTGCACCGTATCA 2489
QY 2341 TTTTGTCTCTCTTGAATCTTGGTGAAGTCTTCTTGAAGAGATCTTCTCATGAT 2400
Db 2490 TTTTGTCTCTCTTGAATCTTGGTGAAGTCTTCTTGAAGAGATCTTCTCATGAT 2549
QY 2401 GATGCGCTTCCAGGAAAAAGACAATTAACCTTGGTTCAGATGAAGAGATCTCTCT 2460
Db 2550 GATGCGCTTCCAGGAAAAAGACAATTAACCTTGGTTCAGATGAAGAGATCTCTCT 2609
QY 2461 GAGACATGCTCAAAATTCAGGCTATTAAATGATGGTTCGATGGCTACTTGGAAATGAA 2520
Db 2610 GAGACATGCTCAAAATTCAGGCTATTAAATGATGGTTCGATGGCTACTTGGAAATGAA 2669
QY 2521 AATTAATCAGTAAATCAGAACTTCTACCTTAAGATTGCTAAACAATATTTGCATAGT 2580
Db 2670 AATTAATCAGTAAATCAGAACTTCTACCTTAAGATTGCTAAACAATATTTGCATAGT 2729
QY 2581 GATGAGACTTGCACAGAAAGGGAATTAAGTAAACCAATATATGTCAGCTTGAAGCTT 2640
Db 2730 GATGAGACTTGCACAGAAAGGGAATTAAGTAAACCAATATATGTCAGCTTGAAGCTT 2789
QY 2641 GCTGCTGGAGTCTATTGGAAGCTGCGACAGAAACCCGTTACCATGTAATCATCACA 2700
Db 2790 GCTGCTGGAGTCTATTGGAAGCTGCGACAGAAACCCGTTACCATGTAATCATCACA 2849
QY 2701 TTAGAACAATATCAGTATGTGCAATTAGCTATCAAGGATGAATGCTATCAAGTAAGCAA 2760
Db 2850 TTAGAACAATATCAGTATGTGCAATTAGCTATCAAGGATGAATGCTATCAAGTAAGCAA 2909
QY 2761 GTGTTGGCCAGAAACTTGCACAAAGGCTTCCCGTTACGCGTTCACACTTGAGTATAG 2820
Db 2910 GTGTTGGCCAGAAACTTGCACAAAGGCTTCCCGTTACGCGTTCACACTTGAGTATAG 2969
QY 2821 GCAATCTGTCCCTTTGTGCAAAAGATCTGTAAAGAGAGAGAGCTCATGCTAGGCAA 2880
Db 2970 GCAATCTGTCCCTTTGTGCAAAAGATCTGTAAAGAGAGAGAGCTCATGCTAGGCAA 3029
QY 2881 TGTTTGGTGAATAATTAATGTAAGGCGGAGTATCTGAAGCAGCATGCAAGCTGTAGT 2940
Db 3030 TGTTTGGTGAATAATTAATGTAAGGCGGAGTATCTGAAGCAGCATGCAAGCTGTAGT 3089
QY 2941 GAAAAATTAATGCTCTCTACAGAGTATGTTCCATATCAATTCACCTTTGGCA 3000
|||||

Db 3090 GAAAAATTAATGCTCTCTTACAGAGTATGTTCTCATATATCAATTCACCTTTGGCA 3149
QY 3001 CATGACCCAGATTAATGTCAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGT 3060
Db 3150 CATGACCCAGATTAATGTCAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGT 3209
QY 3061 CTTTGGTGTCTTGTGGAATATTTAATGGCTAAATAATGAAATTAACAGTCACGCTTTATC 3120
Db 3210 CTTTGGTGTCTTGTGGAATATTTAATGGCTAAATAATGAAATTAACAGTCACGCTTTATC 3269
QY 3121 AGAAAGTGTATATAAATATTAACAACAAAGATGCCAAGACCAAGTATGCAAAA 3180
Db 3270 AGAAAGTGTATATAAATATTAACAACAAAGATGCCAAGACCAAGTATGCAAAA 3329
QY 3181 ATGAATGAANAACCTGTACACTGTGTGATGGTCCATGAATTCATCATGTCAAAAGAT 3240
Db 3330 ATGAATGAANAACCTGTACACTGTGTGATGGTCCATGAATTCATCATGTCAAAAGAT 3389
QY 3241 ACTACATACAGTTTGGAAATCTCTAAAGACCCGGTACTACACAGCTGTTTCTTACTCA 3300
Db 3390 ACTACATACAGTTTGGAAATCTCTAAAGACCCGGTACTACACAGCTGTTTCTTACTCA 3449
QY 3301 CCTGACAAAGATTTAGTAACACCAAAAATTAATGCTCTCTGTAATGAAATCATTTTC 3360
Db 3450 CCTGACAAAGATTTAGTAACACCAAAAATTAATGCTCTCTGTAATGAAATCATTTTC 3509
QY 3361 ACTCTGGAANAACCTTAACCAACCAATGTTCTAGAGAGCTGTAAACAGCACCTTTCATCA 3420
Db 3510 ACTCTGGAANAACCTTAACCAACCAATGTTCTAGAGAGCTGTAAACAGCACCTTTCATCA 3569
QY 3421 GCAGGCAAGCAATCTCAGACCAATATCAGATGAAGAACTGTAAGCAATGCAAGCAGC 3480
Db 3570 GCAGGCAAGCAATCTCAGACCAATATCAGATGAAGAACTGTAAGCAATGCAAGCAGC 3629
QY 3481 AGCTCAAAATCCAAGCTCTCTGGAAGAAATTAAGGGGCGCTTATGTTTGTGAATGAT 3540
Db 3630 AGCTCAAAATCCAAGCTCTCTGGAAGAAATTAAGGGGCGCTTATGTTTGTGAATGAT 3689
QY 3541 CACAGTGAANAATGAAGATTACAAATGCTTTCACCTTTGCGGGGAAAAAAGTGAACA 3600
Db 3690 CACAGTGAANAATGAAGATTACAAATGCTTTCACCTTTGCGGGGAAAAAAGTGAACA 3749
QY 3601 AGAGACGACTCTGATCTTGAAGCTGGAATTTGAGAGAGCTTGAAGCAGGAAAAAAGC 3660
Db 3750 AGAGACGACTCTGATCTTGAAGCTGGAATTTGAGAGAGCTTGAAGCAGGAAAAAAGC 3809
QY 3661 CCCGTCACAGAAAGGAGGAGAAATTAAGTATGATGATGACTTAAGTTGGTGAAGAA 3720
Db 3810 CCCGTCACAGAAAGGAGGAGAAATTAAGTATGATGATGACTTAAGTTGGTGAAGAA 3869
QY 3721 CAGAAACCTTAAGGAGTCAGCAAGTCCGAAAAAGAGGCCATACGGCTTCAGATCTGAT 3780
Db 3870 CAGAAACCTTAAGGAGTCAGCAAGTCCGAAAAAGAGGCCATACGGCTTCAGATCTGAT 3929
QY 3781 GAAACAGAGTGGCTTGAAGGAAAAAGGCTCAAAAGAAATTAATTAAGAAATGAAGTGA 3840
Db 3930 GAAACAGAGTGGCTTGAAGGAAAAAGGCTCAAAAGAAATTAATTAAGAAATGAAGTGA 3989
QY 3841 CAGAAATAGTCCGCAAAAAAGGTTAAAGAGGCGCAACCAAAACCTTGTGGAGGT 3900
Db 3990 CAGAAATAGTCCGCAAAAAAGGTTAAAGAGGCGCAACCAAAACCTTGTGGAGGT 4049
QY 3901 ACACCAAAAGAAAGGCCAACATGAATACTTTTAAAAAAGGAAGCAAAAAAATCTGCA 3960
Db 4050 ACACCAAAAGAAAGGCCAACATGAATACTTTTAAAAAAGGAAGCAAAAAAATCTGCA 4109
QY 3961 CCTCCAGCAGCAGAGAGAGAGAAAGAAAGCAAAAGTGAATACGAAACGAAG 4020
Db 4110 CCTCCAGCAGCAGAGAGAGAGAAAGAAAGCAAAAGTGAATACGAAACGAAG 4169
QY 4021 TCCAAAAGCAAAAGCAGCAGAGTGTCAAGAGAGCAGCAGAGAGCAGAAATCTCTGAA 4080
Db 4170 TCCAAAAGCAAAAGCAGCAGAGTGTCAAGAGAGCAGCAGAGAGCAGAAATCTCTGAA 4229
|||||

OY	4081	TCCTAGTCGAATTGAATCCACACAGCTCCACACACAGAAAGACGACGATCCATCAAAA	4140
Db	4230	TCCTAGTCGAATTGAATCCACACAGCTCCACACACAGAAAGACGACGATCCATCAAAA	4289
OY	4141	ACGCATCACCATCCACACACAAAAAATGTG	4173
Db	4290	ACGCATCACCATCCACACACAAAAAATGTG	4322
RESULT 3			
ID	ABV22430	standard; cDNA: 7473 BP.	
XX	ABV22430;		
AC			
XX			
DT	13-SEP-2002	(first entry)	
XX			
DE	Human prostate expression marker cDNA 22421.		
XX			
KM	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;		
KM	pharmacogenomic marker; gene; ss.		
OS			
XX	Homo sapiens.		
XX			
PN	WO200160860-A2.		
PD			
XX	23-AUG-2001.		
PF			
XX	20-FEB-2001; 2001WO-US05171.		
PR	17-FEB-2000; 2000US-183119P.		
PR	16-MAR-2000; 2000US-189862P.		
PR	23-MAY-2000; 2000US-207454P.		
PR	09-JUN-2000; 2000US-211314P.		
PR	18-JUL-2000; 2000US-219007P.		
PR	13-DEC-2000; 2000US-255281P.		
XX			
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
XX			
PI	Schlegel R, Endege WO, Nonahan JE;		
DR	WPI; 2001-662795/76.		
XX			
PT	Novel isolated nucleic acid molecule associated with cancerous state of		
PT	prostate cells and correlating with presence of prostate cancer, useful		
PT	for detecting presence of prostate cancer, stage of prostate cancer		
XX			
PS	Claim 1; Page 3901-3902; 11750pp; English.		
XX			
CC	The invention relates to an isolated nucleic acid molecule (I) comprising		
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the		
CC	specification or its complement. (I) is useful for:		
CC	(a) assessing whether a patient is afflicted with prostate cancer;		
CC	(b) monitoring the progression of prostate cancer in a patient;		
CC	(c) assessing the efficacy of a test compound to inhibit prostate		
CC	cancer in a patient;		
CC	(d) assessing the efficacy of a therapy for inhibiting prostate cancer		
CC	in a patient;		
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;		
CC	(f) assessing the prostate cell carcinogenic potential of a compound;		
CC	(g) determining whether prostate cancer has metastasized in a patient;		
CC	(h) assessing the aggressiveness or indolence of prostate cancer in a		
CC	patient;		
CC	(I) is also useful as a pharmacodynamic or pharmacogenomic marker.		
XX			
Sequence	7473 BP; 2458 A; 1243 C; 1480 G; 2285 T; 7 other;		
Query Match	99.8%; Score 4165.6; DB 23; Length 7473;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 4168; Conservative	0; Mismatches 4; Indels 0; Gaps 0;		
OY	1	ATGGCTCATTTCAAGACTAGACCAATGATGAAATATTCATATTCGCTGGGCTCAAG	60

Db	158	ATGGCTCATCAAACCTAGACCAATGATGAAAAATTACATATCCGCTGGGTCAAG	217
Qy	61	GAATATCAGATAAAAATATCTAAAGAGAGATGTGAGACGATTTAAAGATGGTGTGAA	120
Db	218	GAATATCAGATAAAAATATCTAAAGAGAGATGCTGACACGATTTAAAGATGGTGTGAA	277
Qy	121	ACTTTATGATATGAGCAGGACTCTAGAGAGAAAAGAGCTTTATTTAACTAGCT	180
Db	278	ACTTTATGATATGAGCAGGACTCTAGAGAGAAAAGAGAGCTTTATTTAACTAGCT	337
Qy	181	TTACATCTGGCTGAGATTTTTTCTCAGACATCCTGGTAAAGATGTGCGTCTCGTA	240
Db	338	TTACATCTGGCTGAGATTTTTTCTCAGACATCCTGTAAAGATGTGCGTCTCGTA	397
Qy	241	GCCTCTCCCTGCTGATATTTTTCAGATTTATGCTCTGGAAGCTCCTTACACATCCCT	300
Db	398	GCCTCTCCCTGCTGATATTTTTCAGATTTATGCTCTGGAAGCTCCTTACACATCCCT	457
Qy	301	GATAAATCAAGAGATATATTTATGTTTATACACAGACGTTGAAGGGCTAGAGATACA	360
Db	458	GATAAATCAAGAGATATATTTATGTTTATACACAGACGTTGAAGGGCTAGAGATACA	517
Qy	361	AAGACCCACATTTCAATAGGATTTTATTACTTGGAACATTGCTGGGTCAAGTCA	420
Db	518	AAGACCCACATTTCAATAGGATTTTATTACTTGGAACATTGCTGGGTCAAGTCA	577
Qy	421	TATACATATGCTTGAGTGAAGATATGACATGAAATTTTACCACACTATACAGAAC	480
Db	578	TATACATATGCTTGAGTGAAGATATGACATGAAATTTTACCACACTATACAGAAC	637
Qy	481	TTATTTTCAGTTATTAACAAATGGCCACATCAGAAAGTCCATATGACANTGCTAGACCT	540
Db	638	TTATTTTCAGTTATTAACAAATGGCCACATCAGAAAGTCCATATGACANTGCTAGACCT	697
Qy	541	ATGAGCTCTATTATTTTGGAAAGTGATACAGTGTCTCAGAGACTTTTGGATACGGTTTA	600
Db	698	ATGAGCTCTATTATTTTGGAAAGTGATACAGTGTCTCAGAGACTTTTGGATACGGTTTA	757
Qy	601	GTAATTCGTGACCGTGCATAGAAATTTAAACAAGCAAGATATGATTTGGCAAAAGCT	660
Db	758	GTAATTCGTGACCGTGCATATGACATTTAAACAAGCAAGATATGATTTGGCAAAAGCT	817
Qy	661	TTACTGAAGAGACAGCTCAAGCTATTTAGGCAATATTTACACACTTTTAACTCAGGTT	720
Db	818	TTACTGAAGAGACAGCTCAAGCTATTTAGGCAATATTTACACACTTTTAACTCAGGTT	877
Qy	721	CTGATGCTTTGGGAAAACATCTATCAGCGATTTGTACAGCATGTCTTGACTTAATTTTG	780
Db	878	CTGATGCTTTGGGAAAACATCTATCAGCGATTTGTACAGCATGTCTTGACTTAATTTTG	937
Qy	781	GAGCTCTCAATATTTGATAGTCATTTGTGTCCTGTCTTTTAAACCACCTTGAATTTTAA	840
Db	938	GAGCTCTCAATATTTGATAGTCATTTGTGTCCTGTCTTTTAAACCACCTTGAATTTTAA	997
Qy	841	TTAAAGAGCAATGATATATGAGAGGCGCTACAGATTTGTTAACTACTGCAAAAATGTTT	900
Db	998	TTAAAGAGCAATGATATATGAGAGGCGCTACAGATTTGTTAACTACTGCAAAAATGTTT	1057
Qy	901	GGGGCAAGATTCAGAAATGGCTTCTCAAAAACAGCCATTGGCAGTGTACTTGGGC	960
Db	1058	GGGGCAAGATTCAGAAATGGCTTCTCAAAAACAGCCATTGGCAGTGTACTTGGGC	1117
Qy	961	AGGTTATATGATATCCATGATCAATCCGCTGGAATGTGGAATTTGCTAGACCAATGT	1020
Db	1118	AGGTTATATGATATCCATGATCAATCCGCTGGAATGTGGAATTTGCTAGACCAATGT	1177
Qy	1021	CTCATGACCAATCCGATTTTGAACAAAACATTTACAGAGATTTCTTAAAGTAGTCAAT	1080
Db	1178	CTCATGACCAATCCGATTTTGAACAAAACATTTACAGAGATTTCTTAAAGTAGTCAAT	1237
Qy	1081	GACCTGAGAGAGCTATTAGACATGATGTTTGTGTCAATATGTTACAGCTGTAAAG	1140

Db 1238 GACCCGTGAGGAGCTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGCTAAAAAG 1297
OY 1141 GATATCTTCTGGTCATGATGACTTACTTATTTTGGAGAGAGAACATTTACACAA 1200
Db 1298 GATATCTTCTGGTCATGATGACTTACTTATTTTGGAGAGAGAACATTTACACAA 1357
OY 1201 CGATGAGAGATACGCAAGAACCCATGATGGACTTGGCCAAATTTATTAAGAAATGCT 1260
Db 1358 CGATGAGAGATACGCAAGAACCCATGATGGACTTGGCCAAATTTATTAAGAAATGCT 1417
OY 1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGATCATAAGACAAATG 1320
Db 1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGATCATAAGACAAATG 1477
OY 1321 CTACATATATATTTATCAAAATAGTATGTGATGATGACTGATGTTGAACGGATCTTGT 1380
Db 1478 CTACATATATATTTATCAAAATAGTATGTGATGATGACTGATGTTGAACGGATCTTGT 1537
OY 1381 CAATACAGTGTCTCAGCAATTTAGAACTACAGAACGGATGAAATGCTTATATTTACTTG 1440
Db 1538 CAATACAGTGTCTCAGCAATTTAGAACTACAGAACGGATGAAATGCTTATATTTACTTG 1597
OY 1441 TATGCGACACTGAGATTTAAATGCTGTGAAGCATTTGAATGAATGGAATGTCAAAAT 1500
Db 1598 TATGCGACACTGAGATTTAAATGCTGTGAAGCATTTGAATGAATGGAATGTCAAAAT 1657
OY 1501 CTGCTCCGACATCAAGTAAGAGATTTGCTGACTGATTAAGCAACCCAAACACATGCT 1560
Db 1658 CTGCTCCGACATCAAGTAAGAGATTTGCTGACTGATTAAGCAACCCAAACACATGCT 1717
OY 1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTAAGAAATTTACCTGATCTGT 1620
Db 1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTAAGAAATTTACCTGATCTGT 1777
OY 1621 AAGGCTCAGGATTTTATGAAGAAATTCACACAGGTGTGAAGATGATGAGAAATTAAGA 1680
Db 1778 AAGGCTCAGGATTTTATGAAGAAATTCACACAGGTGTGAAGATGATGAGAAATTAAGA 1837
OY 1681 AAGCAGTAGAAGTACTGTGTGTCACACATGCTCCTGCAAGCAGGCTGAAGTGTGTG 1740
Db 1838 AAGCAGTAGAAGTACTGTGTGTCACACATGCTCCTGCAAGCAGGCTGAAGTGTGTG 1897
OY 1741 CGTGAATTAAGTAAGAGTTGGGCAACCCCAAGCCTTACAATCTTCTTCTGGAATG 1800
Db 1898 CGTGAATTAAGTAAGAGTTGGGCAACCCCAAGCCTTACAATCTTCTTCTGGAATG 1957
OY 1801 ATCAAGTTCCTTGTGAGAGATAGCACCCTGTGCACATAGATCCGAATCTATCACTGCT 1860
Db 1958 ATCAAGTTCCTTGTGAGAGATAGCACCCTGTGCACATAGATCCGAATCTATCACTGCT 2017
OY 1861 CTTATTTAACAAGTAGAACAAATCAATAGTGAAGCAGAGATGAAGATGAGGTGT 1920
Db 2018 CTTATTTAACAAGTAGAACAAATCAATAGTGAAGCAGAGATGAAGATGAGGTGT 2077
OY 1921 CCAACTGTATCAAGCCATCAGAGCAGTCTTGAATGCTTAAAGTCTCTCATTTACAT 1980
Db 2078 CCAACTGTATCAAGCCATCAGAGCAGTCTTGAATGCTTAAAGTCTCTCATTTACAT 2137
OY 1981 CCCATCTCATTTCTGCTGCAAAATTTGAATCATTAAGTGTGCTGTGAAAATGAT 2040
Db 2138 CCCATCTCATTTCTGCTGCAAAATTTGAATCATTAAGTGTGCTGTGAAAATGAT 2197
OY 2041 GATGAAAAAGTAGCAGAGCTGCACTACAATTTCAAAAAACAGGAAAGCAAAATTTGA 2100
Db 2198 GATGAAAAAGTAGCAGAGCTGCACTACAATTTCAAAAAACAGGAAAGCAAAATTTGA 2257
OY 2101 GAGGATTTTCCACACATCAGATCAGCTTGTCTGTTTACATCACAATCTAAAAA 2160
Db 2258 GAGGATTTTCCACACATCAGATCAGCTTGTCTGTTTACATCACAATCTAAAAA 2317
OY 2161 GGACCCCCCGTCAAGCAAAATATGCAATTCATTTATCATGATGATATTTTCTAGTAA 2220
Db 2318 GGACCCCCCGTCAAGCAAAATATGCAATTCATTTATCATGATGATATTTTCTAGTAA 2377

OY 2221 GAGACCCAGTTTGCACAGATATTTGAGCCTGTGATTAAGAGCCTAGATCAAGCAACCTG 2280
Db 2378 GAGACCCAGTTTGCACAGATATTTGAGCCTGTGATTAAGAGCCTAGATCAAGCAACCTG 2437
OY 2281 GAACATCTCATTAACACATTTGTTACTATTTGGTCAATATTTGCTCTTGCACCTGATCA 2340
Db 2438 GAACATCTCATTAACACATTTGTTACTATTTGGTCAATATTTGCTCTTGCACCTGATCA 2497
OY 2341 TTTGCTGCTCTTGGAAATCTTGGGTAGTACTTTTCATTTGTGAAGATCTTCTCATGAT 2400
Db 2498 TTTGCTGCTCTTGGAAATCTTGGGTAGTACTTTTCATTTGTGAAGATCTTCTCATGAT 2557
OY 2401 GATGCGCTTCCAGGAAAAAGCACTAAACTTTGGGTTCCAGATGAAGAAATCTCTCT 2460
Db 2558 GATGCGCTTCCAGGAAAAAGCACTAAACTTTGGGTTCCAGATGAAGAAATCTCTCT 2617
OY 2461 GAGACAAATGTCAAATTCAGGCTATTAATATGATGGTTGATGGCTACTTGGAAATGA 2520
Db 2618 GAGACAAATGTCAAATTCAGGCTATTAATATGATGGTTGATGGCTACTTGGAAATGA 2677
OY 2521 AATATTCACATTAATCAGGAATCTTACTTAAATGCTTAAACAAATATTTGATAGT 2580
Db 2678 AATATTCACATTAATCAGGAATCTTACTTAAATGCTTAAACAAATATTTGATAGT 2737
OY 2581 GATGAGACTTGCACAGACAGGGAATATTAACACAGATATGTCAGCTGAGACTT 2640
Db 2738 GATGAGACTTGCACAGACAGGGAATATTAACACAGATATGTCAGCTGAGACTT 2797
OY 2641 GCTGCTGGAGTGTATTTGTGAAGCTGCGACAAAGCCCTGTTTACCATGAATATCACA 2700
Db 2798 GCTGCTGGAGTGTATTTGTGAAGCTGCGACAAAGCCCTGTTTACCATGAATATCACA 2857
OY 2701 TTAGAACAATATCAGCTATGTCATTAAGTATCAACAGATGATGCTATCAAGTAAAGCAA 2760
Db 2858 TTAGAACAATATCAGCTATGTCATTAAGTATCAACAGATGATGCTATCAAGTAAAGCAA 2917
OY 2761 GTGTTTGGCCAGAACTTCACAAAGCCCTTCCCTTACGGCTTCCACTGATATATG 2820
Db 2918 GTGTTTGGCCAGAACTTCACAAAGCCCTTCCCTTACGGCTTCCACTGATATATG 2977
OY 2821 GCAATCTGTGCCCTTGTGCAAAAGATCTGTAAGAGAGAGACATCATCTTGAAGCAA 2880
Db 2978 GCAATCTGTGCCCTTGTGCAAAAGATCTGTAAGAGAGAGACATCATCTTGAAGCAA 3037
OY 2881 TGTGTTGGTGAATAATTAATTAAGCGGGAGTATCTGAACACAGATCAGCTGTAT 2940
Db 3038 TGTGTTGGTGAATAATTAATTAAGCGGGAGTATCTGAACACAGATCAGCTGTAT 3097
OY 2941 GAAAAATTAATGTCTCTTCTACAGAGTATGTTGTCATATCAAAATCACTTTGGCA 3000
Db 3098 GAAAAATTAATGTCTCTTCTACAGAGTATGTTGTCATATCAAAATCACTTTGGCA 3157
OY 3001 CATGACCCAGATTAATGTCAAAAGTACAGATATTTGAACAACTTAAAGATGTTAAAGATG 3060
Db 3158 CATGACCCAGATTAATGTCAAAAGTACAGATATTTGAACAACTTAAAGATGTTAAAGATG 3217
OY 3061 CTTTGGTTGTCTGGAATATTAATGGTAAATAATGAATAAAGTACAGCTTTTATC 3120
Db 3218 CTTTGGTTGTCTGGAATATTAATGGTAAATAATGAATAAAGTACAGCTTTTATC 3277
OY 3121 AGAAAGATGATGAATAATTTAAACAACAAAGATGCCCAGAGACAGATATGCAAAA 3180
Db 3278 AGAAAGATGATGAATAATTTAAACAACAAAGATGCCCAGAGACAGATATGCAAAA 3337
OY 3181 ATGAATGAATAACTGTACACTGTGTGTATGTTGCATGAATATCATATGTCAAAGAT 3240
Db 3338 ATGAATGAATAACTGTACACTGTGTGTATGTTGCATGAATATCATATGTCAAAGAT 3397
OY 3241 ACTACATACAGTTTGAATCTCTTAAGAACCCGCTACTACAGCTGCTTCTTCACTCAA 3300
Db 3398 ACTACATACAGTTTGAATCTCTTAAGAACCCGCTACTACAGCTGCTTCTTCACTCAA 3457

QY	3301	CCGTGACAAGAAATTTTCAGTAAACACCAAAAAATTTATCTGCTCCTGAAATGAAATCAATTTTTC	3360
Db	3458	CCTGCACAAGAAATTTTCAGTAAACACCAAAAAATTTATCTGCTCCTGAAATGAAATCAATTTTTC	3517
QY	3361	ACTCCTGGAAAAACCTAAACCAACCAATTTCTAGAGCTGTTAAACAACCCACTTTCATCA	3420
Db	3518	ACTCCTGGAAAAACCTAAACCAACCAATTTCTAGAGCTGTTAAACAACCCACTTTCATCA	3577
QY	3421	GCAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAACCTGTAAGCAATGCAAGC	3480
Db	3578	GCAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAACCTGTAAGCAATGCAAGC	3637
QY	3481	AGCTCAATTCGAAGCTCTCTGGAAGATTAAGGGGAGGCGCTGATGATTTCTGAAATGGAT	3540
Db	3638	AGCTCAATTCGAAGCTCTCTGGAAGATTAAGGGGAGGCGCTGATGATTTCTGAAATGGAT	3697
QY	3541	CACAGTGAATAATGAAAGATTACACAATGCTTCACCTTTGCGGGGGAAAAAAGTGACAG	3600
Db	3698	CACAGTGAATAATGAAAGATTACACAATGCTTCACCTTTGCGGGGGAAAAAAGTGACAG	3757
QY	3601	AGAGACGACTCTGATCTTGTAAAGTCTGAATTTGGAGAAAGCCTTAAGGCGACAGAAAAAAGC	3660
Db	3758	AGAGACGACTCTGATCTTGTAAAGTCTGAATTTGGAGAAAGCCTTAAGGCGACAGAAAAAAGC	3817
QY	3661	CCCGTCACAGAAACAGGAGGAGAAATTTAGTATGATGACTTGACTTAATTTGCTACAGAA	3720
Db	3818	CCCGTCACAGAAACAGGAGGAGAAATTTAGTATGATGACTTGACTTAATTTGCTACAGAA	3877
QY	3721	CAGAAACCTTAAGGCGACTCAGCGAAGTCGAAAAAGAGCCATACGGCTTCAGAAATCTGAT	3780
Db	3878	CAGAAACCTTAAGGCGAGTCAAGCAAGTCGAAAAAGAGCCATACGGCTTCAGAAATCTGAT	3937
QY	3781	GAACAGCGTGGCGCTGAGAGAAAGAGGCTCAAGAAAGATATATTAGAAAAATGAAAGATGAA	3840
Db	3938	GAACAGCGTGGCGCTGAGAGAAAGAGGCTCAAGCAAGATATATTAGAAAAATGAAAGATGAA	3997
QY	3841	CAGAAATAGTCCGCCAAAAAAGGGTAAAGAGGCGCAGCCACCAAAACCTCTTGTTGGAGGT	3900
Db	3998	CAGAAATAGTCCGCCAAAAAAGGGTAAAGAGGCGCAGCCACCAAAACCTCTTGTTGGAGGT	4057
QY	3901	ACACCAAAAGAAAGGACCAACATGAAATCTTCTAAAAAAGAGCAAAAAAATCTGGA	3960
Db	4058	ACACCAAAAGAAAGGACCAACATGAAATCTTCTAAAAAAGAGCAAAAAAATCTGGA	4117
QY	3961	CCTCAGCAGCAGAGGAGGAGGAAGAAAGAAAGACCAAAATGTGAAATACGGAACGAGAG	4020
Db	4118	CCTCAGCAGCAGAGGAGGAGGAAGAAAGAAAGACCAAAATGTGAAATACGGAACGAGAG	4177
QY	4021	TCCAAAAACCAAAAGCACCGAGTGTCAAGAGAGCAGCAGAGGCGAATCTTCCTGAA	4080
Db	4178	TCCAAAAACCAAAAGCACCGAGTGTCAAGAGAGCAGCAGAGGCGAATCTTCCTGAA	4237
QY	4081	TCTAGTGAATTTGAATCTCACACAGTCTCACACACAGAAAGAGACGAGAAACCATCAAAA	4140
Db	4238	TCTAGTGAATTTGAATCTCACACAGTCTCACACACAGAAAGAGACGAGAAACCATCAAAA	4297
QY	4141	ACGCCATTCACATCAACAACCAAAAAAATGT	4172
Db	4298	ACGCCATTCACATCAACAACCAAAAAAATGT	4329
RESULT 4			
ABV25469			
ID	ABV25469	standard; cdna: 7473 BP.	
XX	ABV25469;		
AC			
XX			
XX	16-SEP-2002	(first entry)	
DE			
XX	Human prostate expression marker	cdna 25460.	
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; ss.		

XX	Homo sapiens.
OS	WO200160860-A2.
PN	23 - AUG - 2001.
XX	
PD	20 - FEB - 2001; 2001WO-US05171.
XX	
PP	17 - FEB - 2000; 2000US-183319P.
XX	
PR	16 - MAR - 2000; 2000US-189862P.
PR	25 - MAY - 2000; 2000US-207454P.
PR	09 - JUN - 2000; 2000US-211314P.
PR	18 - JUL - 2000; 2000US-219007P.
PR	13 - DEC - 2000; 2000US-25281P.
XX	
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI	Schlegel R, Endege WO, Monahan JE;
XX	
DR	WPI: 2001-662795/76.
XX	
PT	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -
PT	
PS	Claim 1; Page 5055-5057; 11750pp; English.
XX	
CC	The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:
CC	(a) assessing whether a patient is afflicted with prostate cancer;
CC	(b) monitoring the progression of prostate cancer in a patient;
CC	(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
CC	(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;
CC	(f) assessing the prostate cell carcinogenic potential of a compound;
CC	(g) determining whether prostate cancer has metastasized in a patient;
CC	(h) assessing the aggressiveness or indolence of prostate cancer in a patient;
CC	(I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX	
SQ	Sequence 7473 BP; 2457 A; 1243 C; 1481 G; 2285 T; 7 other;
Query Match	. 99.8%; Score 4164; DB 23; Length 7473;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 4167; Conservative	0; Mismatches 5; Indels 0; Gaps 0
OY	1 ATGGCTCATTCGCAAGACTAGACCAATGATGAAATAATTACATATCCGCCTGGGGTCAAG 60 DB 158 ATGGCTCATTCGAAGACTAGACCAATGATGAAATAATTACATATCCGCCTGGGGTCAAG 217 OY 61 GAATATACAGATTAATAATATCTAAAGAAGAGATGTGACAGCATTTAAGAAGCGTTGCCAAA 120 DB 218 GAAATATACAGATTAATAATATCTAAAGAAGAGATGTGACAGCATTTAAGAAGCGTTGCCAAA 277 OY 121 ACTTTTATGATATGACACGAGACTCTGAAGAAGAAAGAGAGCTTTATTTAAACCTAGCT 180 DB 278 ACTTTTATGATATGACACGAGACTCTGAAGAAGAAAGAGAGCTTTATTTAAACCTAGCT 337 OY 181 TTACATCTTGCTTCAGATTTTTTTTCTCANGCATCTCGGTAAAGATGTTGCTTACCTGCTA 240 DB 338 TTACATCTTGCTTCAGATTTTTTTTCTCANGCATCTCGGTAAAGATGTTGCTTACCTGCTA 397 OY 241 GCCGCGCGCTTCGCTGATATTTTTCAGAGATTTATGCTCTGAAAGCGCTTTAACATCCCCT 300 DB 398 GCCGCGCGCTTCGCTGATATTTTTCAGAGATTTATGCTCTGAAAGCGCTTTAACATCCCCT 457 OY 301 GATTAACCTAAAGATATATTTATGTTTATTAACAACAAGCTGAAGGGCTGAGAGATACA 360 DB 458 GATTAACCTAAAGATATATTTATGTTTATTAACAACAAGCTGAAGGGCTGAGAGATACA 517

QY	361	AAGAGCCACAAATTCATAGTATTTTATTTACTTGAGAACATTGCTTGGCTCAAGTCA	420
Db	518	AAGAGCCACCAATTCATAGGATATTTTATTACTTGAGAACATTGCTTGGCTCAAGTCA	577
QY	421	TATACATATCTTGGAGTTAGAGATATGCAATGAATTTTCACCAGCTATACGAAC	480
QY	481	TTATTTTGTAGTTTAAACAATGGCCACATCGAAGAGTCACATATGACATGGTAAAGCT	540
Db	638	TTATTTTGTAGTTTAAACAATGGCCACATCGAAGAGTCACATATGACATGGTAAAGCT	697
QY	541	ATGAGCTCTATTATTGTGAGAGGTATACAGTGTCTCAGAGACTTTTGGATACGGTTTGA	600
Db	698	ATGAGCTCTATTATTGTGAGAGGTATACAGTGTCTCAGAGACTTTTGGATACGGTTTGA	757
QY	601	GTAATATCTGTACCTGCTCATTAAGAATTTAAACAAGCAAGCATATGATTTTGGCAAGCT	660
Db	758	GTAATATCTGTACCTGCTCATTAAGAATTTAAACAAGCAAGCATATGATTTTGGCAAGCT	817
QY	661	TTACTGAGAGAGACGCTCAAGCTTTGAGCCATATTTACACATTTTATATACAGT	720
Db	818	TTACTGAGAGAGACGCTCAAGCTTTGAGCCATATATTTACCAATTTTATATACAGT	877
QY	721	CTGATGCTTGGGAAAAACATCTATCAGCAGATTGTGCAGAGCATGTCTTTGACTTAAATTG	780
Db	878	CTGATGCTTGGGAAAAACATCTATCAGCAGATTGTGCAGAGCATGTCTTTGACTTAAATTG	937
QY	781	GAGCTCTTCAATATTGATAGTCATTGTGCTGCTCTCTGTGTTTACCCAGCTTGCAATTTTAA	840
Db	938	GAGCTCTTCAATATTGATAGTCATTGTGCTGCTCTCTGTGTTTACCCAGCTTGCAATTTTAA	997
QY	841	TTAAAGACCAATGATTAATGAGAGGCGCTACAAGTGTGTTAACTCTGGCAAAATGT	900
Db	998	TTAAAGACCAATGATTAATGAGAGGCGCTACAAGTGTGTTAACTCTGGCAAAATGT	1055
QY	901	GGGGCAAAAGATTGCAATTTGGCTTCTCAAAACAAGCCACTTTGGCAGTGTACTTGGGC	960
Db	1058	GGGGCAAAAGATTGCAATTTGGCTTCTCAAAACAAGCCACTTTGGCAGTGTACTTGGGC	1113
QY	961	AGGTTTAAATGATATCCATGTACCAATCCGCGCTGAATGTGTGAAATTTGCTAGCCATTGT	1020
Db	1118	AGGTTTAAATGATATCCATGTACCAATCCGCGCTGAATGTGTGAAATTTGCTAGCCATTGT	1177
QY	1021	CTCATGAACCAATCCGATTTTAGCAAAACAATTACAGATATCTTAAAGTGAGTCAAT	1088
Db	1178	CTCATGAACCAATCCGATTTAGCAAAACAATTACAGATATCTTAAAGTGAGTCAAT	1233
QY	1081	GACCCGTGAGAGCAATATAGACATATGTTTGTGTCAATATGATACAGCTCTAAAG	1140
Db	1238	GACCCGTGAGAGCAATATAGACATATGTTTGTGTCAATATGATACAGCTCTAAAG	1293
QY	1141	GATATTTCTTGGTCAATGATCTACTTAAATTTGTGAGAGAGAACTTTAGACAAA	1200
Db	1298	GATATTTCTTGGTCAATGATCTACTTAAATTTGTGAGAGAGAACTTTAGACAAA	1355
QY	1201	CGATGAGAGATACGAAAGAAAGCCATGATGGGACTTGCCCAATTTATTAAGAAATATGCT	1266
Db	1358	CGATGAGAGATACGAAAGAAAGCCATGATGGGACTTGCCCAATTTATTAAGAAATATGCT	1413
QY	1261	TTACAGTACGAGCTGGAAAGATGCTCCAAACAAGATAGAGATGATCAAAAGCAAAATG	1332
Db	1418	TTACAGTACGAGCTGGAAAGATGCTCCAAACAAGATAGAGATGATCAAAAGCAAAATG	1477
QY	1321	CTACATATATATTTATCAAAATAGTATGATGATGACATCTTGTGGAACGATCTTTGCT	1380
Db	1478	CTACATATATATTTATCAAAATAGTATGATGATGACATCTTGTGGAACGATCTTTGCT	1533
QY	1381	CAATATCATGTTCTCTCAACAATTTAGAAACTACAGACGAGTGAATGCTTATATATCTTG	1440
Db	1538	CAATATCATGTTCTCTCAACAATTTAGAAACTACAGACGAGTGAATGCTTATATATCTTG	1597

QY	1441	TATGCCACACGTGATTTTAAATGCTGTGAACACATTGGAATGAAATGTCAAAAT	1500
Db	1598	TATGCCACACGTGATTTTAAATGCTGTGAACACATTGGAATGAAATGTCAAAAT	1657
QY	1501	CTGCCTCCGACATCAAGTAAAGAGATTGGCTTCACTTGATTAAAGCAACCCAAACAGATGCC	1560
Db	1658	CTGCCTCCGACATCAAGTAAAGAGATTGGCTTCACTTGATTAAAGCAACCCAAACAGATGCC	1717
QY	1561	AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAAGAAATTTTACCTGATCTGGT	1620
Db	1718	AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAAGAAATTTTACCTGATCTGGT	1777
QY	1621	AAGGCTCAGGATTTTCATATGACAAATTCACACAGCTGTTAGAAATGATGAGAAATAAGA	1680
Db	1778	AAGGCTCAGGATTTTCATATGACAAATTCACACAGCTGTTAGAAATGATGAGAAATAAGA	1837
QY	1681	AAGCAGTTAGAAAGTACTTGTAGTCCAAATGCTCCCTCAGCAGGGCTGAAGGTTGGTGTG	1740
Db	1838	AAGCAGTTAGAAAGTACTTGTAGTCCAAATGCTCCCTCAGCAGGGCTGAAGGTTGGTGTG	1897
QY	1741	CGTGAATTAACATAAGAAAGTTGGGCAACCCCAACAGCCTACAAATCTCTTCTGGAAATG	1800
Db	1898	CGTGAATTAACATAAGAAAGTTGGGCAACCCCAACAGCCTACAAATCTCTTCTGGAAATG	1957
QY	1801	ATCAAGTTTCTCTTGGAGAGATACACACCTTGCCACATAGATACCGAATCTATCAGTCT	1860
Db	1958	ATCAAGTTTCTCTTGGAGAGATAGCACCTTGCCACATAGATACCGAATCTATCAGTCT	2017
QY	1861	CTTATTAAACAAAGTGAACAAATTCATATGATGGAACAGATGATGAAATGAGAGTCTT	1920
Db	2018	CTTATTAAACAAAGTGAACAAATTCATATGATGGAACAGATGATGAAATGAGAGTCTT	2077
QY	1921	CCAACGTATCAAGCCATCAGACAGAGCTTTGAACGCTTAAAGTACTCTCATTTACACAT	1980
Db	2078	CCAACGTATCAAGCCATCAGACAGAGCTTTGAACGCTTAAAGTACTCTCATTTACACAT	2137
QY	1981	CCCATCTCATTTTCATCTGCTGAACAACTTTGAATCATATTCAGCTTGTCTGAAATGGAT	2040
Db	2138	CCCATCTCATTTTCATCTGCTGAACAACTTTGAATCATATTCAGCTTGTCTGAAATGGAT	2197
QY	2041	GATGAAAAAAGTAGCAGAAAGCTGCACATACAAATTTTCAAAAACACAGAACCAAAATTGAA	2100
Db	2198	GATGAAAAAAGTAGCAGAAAGCTGCACATACAAATTTTCAAAAACACAGAACCAAAATTGAA	2257
QY	2101	GAGATTTTTCACACATACAGATACAGCCCTGCTCCTGTTTACATCCAAATCTTAAAAAA	2160
Db	2258	GAGATTTTTCACACATACAGATACAGCCCTGCTCCTGTTTACATCCAAATCTTAAAAAA	2317
QY	2161	GGACCCCCCGCTCAAGCCAAATATGSCATTCATTTGATATCCATGGGATATTTTCTAGTAAA	2220
Db	2318	GGACCCCCCGCTCAAGCCAAATATGSCATTCATTTGATATCCATGGGATATTTTCTAGTAAA	2377
QY	2221	GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATTAAGAGCTTAGATCCAAAGACCTG	2280
Db	2378	GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATTAAGAGCTTAGATCCAAAGACCTG	2437
QY	2281	GAACATCTCATAAACACCAATGCTGTTACATATGGTCATATGCTCTCCCTGACACCTGATCAA	2340
Db	2438	GAACATCTCATAAACACCAATGCTGTTACATATGGTCATATGCTCTCCCTGACACCTGATCAA	2497
QY	2341	TTTGTGCTGCTTGGAAATCTTTGGGTAGGCTACTTTCATTTGAAAGATCTTCTCATGAT	2400
Db	2498	TTTGTGCTGCTTGGAAATCTTTGGGTAGGCTACTTTCATTTGAAAGATCTTCTCATGAT	2557
QY	2401	GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAATCTCTCT	2460
Db	2558	GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAATCTCTCTCT	2617
QY	2461	GAGACAAATGCTCAAAATTCACAGCTATTAAAAATGATGGTTCGATGGCTACTTGGAAATGAAA	2520
Db	2618	GAGACAAATGCTCAAAATTCACAGCTATTAAAAATGATGGTTCGATGGCTACTTGGAAATGAAA	2677
QY	2521	AATTAATCAGAGTAATCAGAAACTTCACTTACCTTAGATTTGCTTAAACAACTTTTGCATAGT	2580

```

Db 2678 AATATCAGCTAATCAGGAACCTTACCTTAAGATGTACAAACAATATGTGATAGT 2737
Oy 2581 GATGAGAGCTTGACAGGAAGGGGAAAAATTAGTAACAGATATGTCACGCTGAGACT 2640
Db 2738 GATGAGAGCTTGACAGGAAGGGGAAAAATTAGTAACAGATATGTCACGCTGAGACT 2797
Oy 2641 GCTGCTGGAGTGTATGTAGAGCTGACAGAACCCCTGTACATGAATCAATCACA 2700
Db 2798 GCTGCTGGAGTGTATGTAGAGCTGACAGAACCCCTGTACATGAATCAATCACA 2857
Oy 2701 TTAACAACATATCAGCTATGTGATTTAGCTATCAACGATGTAATGCTATCAAGTAA 2760
Db 2858 TTAACAACATATCAGCTATGTGATTTAGCTATCAACGATGTAATGCTATCAAGTAA 2917
Oy 2761 GTGTTGGCCAGAACTTCACAAAGGCTTCCTCCGTTAGCGCTCCACTGAGATATG 2820
Db 2918 GTGTTGGCCAGAACTTCACAAAGGCTTCCTCCGTTAGCGCTTCCTGAGATATG 2977
Oy 2821 GCAATCTGTGCCCTTGTGCAAAAGATCCTGTAAAGAGAGAGAGAGATGCTAGGCA 2880
Db 2978 GCAATCTGTGCCCTTGTGCAAAAGATCCTGTAAAGAGAGAGAGATGCTAGGCA 3037
Oy 2881 TGTGTTGGAAAAATATTAATGTAAAGCGGAGATCTGAAAGCAGATGCAAGCTGTAGT 2940
Db 3038 TGTGTTGGAAAAATATTAATGTAAAGCGGAGATCTGAAAGCAGATGCAAGCTGTAGT 3097
Oy 2941 GAAAAATATATGCTCTCTACAGAGATGTTGTCATATACAAATCACCTTTGGCA 3000
Db 3098 GAAAAATATATGCTCTCTCTACAGAGATGTTGTCATATACAAATCACCTTTGGCA 3157
Oy 3001 CATGACCCAGATATATGTCAAAGTACAGAGATTTGAAACAATTAAAGATTTAAAGATGT 3060
Db 3158 CATGACCCAGATATATGTCAAAGTACAGAGATTTGAAACAATTAAAGATTTAAAGATGT 3217
Oy 3061 CTTGTTGTTGTTGGAATATTAATGCTGTAAGTAAATGAATACAGTACAGCTTTATC 3120
Db 3218 CTTGTTGTTGTTGGAATATTAATGCTGTAAGTAAATGAATACAGTACAGCTTTATC 3277
Oy 3121 AGAAGATGTAGAAAAATATTAACAAACAAGATGCGCCAGAGACAGATGATGCAAAA 3180
Db 3278 AGAAGATGTAGAAAAATATTAACAAACAAGATGCGCCAGAGACAGATGATGCAAAA 3337
Oy 3181 ATGATGAAAAAAGTACACTGTGTGATGTGTCATGAATATCATCATGTCAAAGAGT 3240
Db 3338 ATGATGAAAAAAGTACACTGTGTGATGTGTCATGAATATCATCATGTCAAAGAGT 3397
Oy 3241 ACTCATATCAGTTTGGAAATCTCTAAAGACCGGACTACAGAGCTGTTCTCTCACTGA 3300
Db 3398 ACTCATATCAGTTTGGAAATCTCTCTAAAGACCGGACTACAGAGCTGTTCTCTCACTGA 3457
Oy 3301 CCTGACAAAGATTTTCAGTAAACCAAAATATATGCTCCTCGAATGAATCATTTTTC 3360
Db 3458 CCTGACAAAGATTTTCAGTAAACCAAAATATATGCTCCTCGAATGAATCATTTTTC 3517
Oy 3361 ACTCCTGGAAAACTTAAACAACCAATGTTCTAGAGCTGTTAACAGCCACTTTCATCA 3420
Db 3518 ACTCCTGGAAAACTTAAACAACCAATGTTCTAGAGCTGTTAACAGCCACTTTCATCA 3577
Oy 3421 GCAGGCAAGCAATCTCAGCAATCATCAGAAATGGAATGTAAGAAATGCAAGCAGC 3480
Db 3578 GCAGGCAAGCAATCTCAGCAATCATCAGAAATGGAATGTAAGAAATGCAAGCAGC 3637
Oy 3481 AGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGGGAGGCTTGATAGTTCTGAAATGAT 3540
Db 3638 AGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGGGAGGCTTGATAGTTCTGAAATGAT 3697
Oy 3541 CACAGTGAATAATGAAGATTACACAATGTCTTACCTTCCGCGGAAAAAAGTACAG 3600
Db 3698 CACAGTGAATAATGAAGATTACACAATGTCTTACCTTCCGCGGAAAAAAGTACAG 3757
Oy 3601 AGAAGCACTCTGATCTTGTAGAGCTGAATTTGGAAGCCTAGAGGAGGAAAAAAGC 3660

```

```

Db 3758 AGAGAGCACTCTGATCTTGTAAAGCTCTGAATTTGGAAGACCTAGAGCAGGAAAAAAGC 3817
Oy 3661 CCCGTCACAGAAAGAGAGAGAAATTAGTATGATGACTGTACATAAGTTGGTACAGAA 3720
Db 3818 CCCGTCACAGAAAGAGAGAGAAATTAGTATGATGACTGTACATAAGTTGGTACAGAA 3877
Oy 3721 CAGAAACCTTAAGGAGTACAGAGAAAGTGGGAAAAAGAGCCATACGGCTTCAGAAATCTGAT 3780
Db 3878 CAGAAACCTTAAGGAGTACAGAGAAAGTGGGAAAAAGAGCCATACGGCTTCAGAAATCTGAT 3937
Oy 3781 GAAACGAGTGGCTTGAGGAAAAAGGCTCAAGAGATATATTGAAAAATGAGATGAA 3840
Db 3938 GAAACGAGTGGCTTGAGGAAAAAGGCTCAAGAGATATATTGAAAAATGAGATGAA 3997
Oy 3841 CAGAAATAGTCCGCCCCAAAAAAGGTTAAAGAGGCGCACACCAAAACCTTGGTGGAGGT 3900
Db 3998 CAGAAATAGTCCGCCCCAAAAAAGGTTAAAGAGGCGCACACCAAAACCTTGGTGGAGGT 4057
Oy 3901 ACACCAAAAGAAAGACCAATGAAACTTCTAATAAAGAAAGCAAAAAAATCTGGA 3960
Db 4058 ACACCAAAAGAAAGACCAATGAAACTTCTAATAAAGAAAGCAAAAAAATCTGGA 4117
Oy 3961 CCTCAGCACCAGAGAGAGAGAGAGAAAGACAAAGTGAATTCGAAACAGAG 4020
Db 4118 CCTCAGCACCAGAGAGAGAGAGAGAAAGACAAAGTGAATTCGAAACAGAG 4177
Oy 4021 TCCAAAAGCAAAAGACAGCAGAGTGTCAAGAGAGACAGCAGAGAGCAATCTCTGAA 4080
Db 4178 TCCAAAAGCAAAAGACAGCAGAGTGTCAAGAGAGACAGCAGAGAGCAATCTCTGAA 4237
Oy 4081 TCTAGTGAATTTGAATTCACACAGTCCACACCAGAAAGAGAGAGCAATCAAAA 4140
Db 4238 TCTAGTGAATTTGAATTCACACAGTCCACACCAGAAAGAGAGAGCAATCAAAA 4297
Oy 4141 AGCCATCAGCATCCACAAACCAAAAAAATATG 4172
Db 4298 AGCCATCAGCATCCACAAACCAAAAAAATATG 4329

RESULT 5
ABV28244
ID ABV28244 standard; cDNA; 7473 BP.
XX
AC ABV28244;
XX
DE 16-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 28235.
XX
KW Human; prostate cancer; cytosolic; carcinogen; pharmacodynamic marker;
XX
OS pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX
XX 16-MAR-2000; 2000US-189862P.
XX
XX 25-MAY-2000; 2000US-207454P.
XX
XX 09-JUN-2000; 2000US-211314P.
XX
XX 18-JUL-2000; 2000US-219007P.
XX
XX 13-DEC-2000; 2000US-235281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI: 2001-662795/76.
XX

```

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1: Page 5869-5870; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 7473 BP; 2458 A; 1244 C; 1479 G; 2285 T; 7 other:

Query Match 99.8%; Score 4164; DB 23; Length 7473;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 AAGGCTCATCAAAAGCTAGGACCAATGATGAGAAAATACATATCCGCTGGGGTCAAG 60
DB 158 AAGGCTCATCAAAAGCTAGGACCAATGATGAGAAAATACATATCCGCTGGGGTCAAG 217

OY 61 GAAATATCGATTAATATCTAAGAGAGATGTCAGACGATTAAGATGCTTGTGAAA 120
DB 218 GAAATATCGATTAATATCTAAGAGAGATGTCAGACGATTAAGATGCTTGTGAAA 277

OY 121 ACTTTATGATATGAGACGACGACTGTGAAGAGAAAGAGCTTATTTAACTAGCT 180
DB 278 ACTTTATGATATGAGACGACGACTGTGAAGAGAAAGAGCTTATTTAACTAGCT 337

OY 181 TTACATCTTGCCTTCAGATTTTTCCTCAAGCATCCGCTGAAGATGTTGCTTACTGTA 240
DB 338 TTACATCTTGCCTTCAGATTTTTCCTCAAGCATCCGCTGAAGATGTTGCTTACTGTA 397

OY 241 GCCTGCTGCTGCTGATATTTTCAGATTTTATGCTGCAAGCTCCTTACACATCCCT 300
DB 398 GCCTGCTGCTGCTGATATTTTCAGATTTTATGCTGCAAGCTCCTTACACATCCCT 457

OY 301 GATTAACCTAAGGATATATTTATGTTTATACAAAGACAGTGAAGGGCTAGAGATAGA 360
DB 458 GATTAACCTAAGGATATATTTATGTTTATACAAAGACAGTGAAGGGCTAGAGATAGA 517

OY 361 AAGAGCCCAATTCATAGATTTTATTTTACTTGAGAACATTTGCTGGCTCAAGTCA 420
DB 518 AAGAGCCCAATTCATAGATTTTATTTTACTTGAGAACATTTGCTGGCTCAAGTCA 577

OY 421 TATTAACATATGCTTGAGATTAAGATAGCAATGAATTTTCCAGCAGTATCAGAAC 480
DB 578 TATTAACATATGCTTGAGATTAAGATAGCAATGAATTTTCCAGCAGTATCAGAAC 637

OY 481 TATTTTTCAGTTATTAACCAATGGCCCAATCAGAAAGTCCATATGACATGCTAGACCTT 540
DB 638 TATTTTTCAGTTATTAACCAATGGCCCAATCAGAAAGTCCATATGACATGCTAGACCTT 697

OY 541 AAGACCTCATATTTTGAAGGTGATAGAGTCTCAGAGAGCTTTGGATACGTTTAA 600
DB 698 AAGACCTCATATTTTGAAGGTGATAGAGTCTCAGAGAGCTTTGGATACGTTTAA 757

OY 601 GTAATCTGCTACTGCTCATAGAATTTAAACAAGCAGCATATGATTTGGCAAGGCT 660
DB 758 GTAATCTGCTACTGCTCATAGAATTTAAACAAGCAGCATATGATTTGGCAAGGCT 817

OY 661 TTACTGAAGAGACAGCTCAAGCTATGAGCATATATTTACCACTTTTATATCAGGTT 720

DB 818 TTACTGAAGAGACAGCTCAAGCTATGAGCCATATATTTACCAATTTTATATCAGGTT 877

OY 721 CTGATCTGTGGAAAACATCTATACAGCATTTGTCAGAGATGCTTTGACTTAATTTTG 780

DB 878 CTGATCTGTGGAAAACATCTATACAGCATTTTGTACAGAGATGCTTTGACTTAATTTTG 937

OY 781 GAGCTCTACAAATTTATATAGTCAATTTGCTGCTGCTTTTACCAGCTTGAATTTAA 840

DB 938 GAGCTCTACAAATTTATATAGTCAATTTGCTGCTGCTTTTACCAGCTTGAATTTAA 997

OY 841 TTAAGAGCAATGATATAGAGAGCGCTTACAAAGTTTAACTACTGCGCAAAATGTTT 900

DB 998 TTAAGAGCAATGATATATAGAGAGCGCTTACAAAGTTTAACTACTGCGCAAAATGTTT 1057

OY 901 GGGGCAAGAGATTCAGAAATTTGGCTTCAAAACACAGCACTTTGGCAGTCTACTGGGC 960

DB 1058 GGGGCAAGAGATTCAGAAATTTGGCTTCAAAACACAGCACTTTGGCAGTCTACTGGGC 1117

OY 961 AAGTTTAATGATATCCATGATACCAATCCGCTGGAATGTGTAAATTTGCTAGCCATTTG 1020

DB 1118 AAGTTTAATGATATCCATGATACCAATCCGCTGGAATGTGTAAATTTGCTAGCCATTTG 1177

OY 1021 CTCATGAACCATCTGATTTTACGAAAGACTTAAAGAGATATCTTAAAGTGAAGTCAAT 1080

DB 1178 CTCATGAACCATCTGATTTTACGAAAGACTTAAAGAGATATCTTAAAGTGAAGTCAAT 1237

OY 1081 GACCTTGAGAGAGCTTTTACAGATGATTTTGTCTCAATAGTTCACAGCTCTAAAG 1140

DB 1238 GACCTTGAGAGAGCTTTTACAGATGATTTTGTCTCAATAGTTCACAGCTCTAAAG 1297

OY 1141 GATATCTTCTGCTCAATGATCCTACTTAATTTTGTGAGAGAGAAACATTAAGCAAA 1200

DB 1298 GATATCTTCTGCTCAATGATCCTACTTAATTTTGTGAGAGAGAAACATTAAGCAAA 1357

OY 1201 CGATGAGAGATAGCAAAAGAGCCATGATGGAATTTGCCAAATTTATTAAGAAATATGCT 1260

DB 1358 CGATGAGAGATAGCAAAAGAGCCATGATGGAATTTGCCAAATTTATTAAGAAATATGCT 1417

OY 1261 TTACAGTCAAGAGCTGGAAGAGTGTGCAAAACACATAGCATGATCAAAAGCAATTTG 1320

DB 1418 TTACAGTCAAGAGCTGGAAGAGTGTGCAAAACACATAGCATGATCAAAAGCAATTTG 1477

OY 1321 CTACATATATATATATATAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1380

DB 1478 CTACATATATATATATAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1537

OY 1381 CAATACATGCTGCTCACAATTTAGAACTACAGAAACGATGAATGCTTATATTTACTTG 1440

DB 1538 CAATACATGCTGCTCACAATTTAGAACTACAGAAACGATGAATGCTTATATTTACTTG 1597

OY 1441 TATGCCACACTGGATTTAAATGCTGGAAGAGCTTGAAGTAATGTGGAATGTCAAAAT 1500

DB 1598 TATGCCACACTGGATTTAAATGCTGGAAGAGCTTGAAGTAATGTGGAATGTCAAAAT 1657

OY 1501 CTGCTCCGACATCAAGTAAAGATTTGCTTGAATTTTAAAGCAACCCAAACAGATGCTC 1560

DB 1658 CTGCTCCGACATCAAGTAAAGATTTGCTTGAATTTTAAAGCAACCCAAACAGATGCTC 1717

OY 1561 AGTGTCAAGGCTATATTTTCAAAAGTGTATTTATTAAGAAATTTTACCTGATCTGCT 1620

DB 1718 AGTGTCAAGGCTATATTTTCAAAAGTGTATTTATTAAGAAATTTTACCTGATCTGCT 1777

OY 1621 AAGGCTCAGAGATTTTCAAGAAATTTCAACAGAGTGTGAAGATATATGAAGAAATTA 1680

DB 1778 AAGGCTCAGAGATTTTCAAGAAATTTCAACAGAGTGTGAAGATATATGAAGAAATTA 1837

OY 1681 AAGCATGTGAAGTACTTGTATGTCACAAATCTCTGCAAGAGCTCAAGGTTGTGTG 1740

DB 1838 AAGCATGTGAAGTACTTGTATGTCACAAATCTCTGCAAGAGCTCAAGGTTGTGTG 1897

OY 1741 CGTGAATTAAGTAAAGTTGGCAACCCCAACAGCTTACAAATCTTTCTGGAATG 1800

Db 1898 CGTGAATACTAAGAGTTGGGCAACCCCAAGCCTACAAATCCTTTCCTGGAAATG 1957
Qy 1801 ATCAAGTTCTCTTGAGAGATAGACCTGTGCACATAGATACCAATCTATCAGTGT 1860
Db 1958 ATCAAGTTCTCTTGAGAGATAGACCTGTGCACATAGATACCAATCTATCAGTGT 2017
Qy 1861 CTATTTAAACAAGTGAACAAATCAATAGATGAACAGAGATGAAGATGAGGTGT 1920
Db 2018 CTATTTAAACAAGTGAACAAATCAATAGATGAACAGAGATGAAGATGAGGTGT 2077
Qy 1921 CCAACTGATCAAGCATAGAGACAGGCTTGAACCTGCTTAAGTCTCTCATTTACAT 1980
Db 2078 CCAACTGATCAAGCATAGAGACAGGCTTGAACCTGCTTAAGTCTCTCATTTACAT 2137
Qy 1981 CCAATCTCATTTTCATTTCTGTGAAACATTTGATCATTTACTGGCTGTGCTGAAATGAT 2040
Db 2138 CCAATCTCATTTTCATTTCTGTGAAACATTTGATCATTTACTGGCTGTGCTGAAATGAT 2197
Qy 2041 GATGAAAAAGTAGCAAGAGCTGCATACAAATTTTCAAAAAACACAGGAAGCAAAATGAA 2100
Db 2198 GATGAAAAAGTAGCAAGAGCTGCATACAAATTTTCAAAAAACACAGGAAGCAAAATGAA 2257
Qy 2101 GAGGTTTTCCACATCAGATCAGGCTTGTCTCTGTTTACATCACAATCTAAAAA 2160
Db 2258 GAGGTTTTCCACATCAGATCAGGCTTGTCTCTGTTTACATCACAATCTAAAAA 2317
Qy 2161 GGACCCCCCGTCAAGCCAAATATGCAATTCATTTGATCATGCGATATTTTCTAGTAA 2220
Db 2318 GGACCCCCCGTCAAGCCAAATATGCAATTCATTTGATCATGCGATATTTTCTAGTAA 2377
Qy 2221 GAGACCCAGTTTGCACAGATTTTGAAGCTTGCATTAAGAGCTTAGATCCAAAGCAACGTG 2280
Db 2378 GAGACCCAGTTTGCACAGATTTTGAAGCTTGCATTAAGAGCTTAGATCCAAAGCAACGTG 2437
Qy 2281 GAACATCTCATACACCATGTTTACTATTTGGTCTATTTGGCTCTCTTGCACCTGATCAA 2340
Db 2438 GAACATCTCATACACCATGTTTACTATTTGGTCTATTTGGCTCTCTTGCACCTGATCAA 2497
Qy 2341 TTTGCTGCTCTTGGAAATCTTGGTAGCTACTTTCATTTGAAAGATCTTCTCATGAT 2400
Db 2498 TTTGCTGCTCTTGGAAATCTTGGTAGCTACTTTCATTTGAAAGATCTTCTCATGAT 2557
Qy 2401 GATGGCTTCCAGGAAAAAAGACAATTAACCTTTGGTTCCAGATGAAGAGTATCTCT 2460
Db 2558 GATGGCTTCCAGGAAAAAAGACAATTAACCTTTGGTTCCAGATGAAGAGTATCTCT 2617
Qy 2461 GAGACAAATGTCAAAAATCAGGCTATTAATAATGATGGTGGCTACTTGGAAATGAAA 2520
Db 2618 GAGACAAATGTCAAAAATCAGGCTATTAATAATGATGGTGGCTACTTGGAAATGAAA 2677
Qy 2521 AATATCACAAGTAAATCAGAACTTCTACTTAAGATGCTAACAAATATTCATAGT 2580
Db 2678 AATATCACAAGTAAATCAGAACTTCTACTTAAGATGCTAACAAATATTCATAGT 2737
Qy 2581 GATGAGACTTGAAGAAAGAGGGAATTAAGTAAACAGATATGTCACGCTGAGACTT 2640
Db 2738 GATGAGACTTGAAGAAAGAGGGAATTAAGTAAACAGATATGTCACGCTGAGACTT 2797
Qy 2641 GCTGCTGGAGTGGCATGTGAAGCTGCAAGAACCTGTTNCCAGTAAATCTTCACA 2700
Db 2798 GCTGCTGGAGTGGCATGTGAAGCTGCAAGAACCTGTTNCCAGTAAATCTTCACA 2857
Qy 2701 TTGAACAAATATCAGCTATGTGATTAAGTATCAACAGATGATGCTATCAAGTAAAGAA 2760
Db 2858 TTGAACAAATATCAGCTATGTGATTAAGTATCAACAGATGATGCTATCAAGTAAAGAA 2917
Qy 2761 GTGTTTGGCCAGAACTTCAACAAAGGCTTTCGGTTTACGGCTTCCACTTGAGTATATG 2820
Db 2918 GTGTTTGGCCAGAACTTCAACAAAGGCTTTCGGTTTACGGCTTCCACTTGAGTATATG 2977
Qy 2821 GCATCTGTGCTTTTGTGCAAAAGATCTGTAAAGAGAGAAAGCTCATGCTAGGCAA 2880
Db 2978 GCATCTGTGCTTTTGTGCAAAAGATCTGTAAAGAGAGAGCTCATGCTAGGCAA 3037

Qy 2881 TGTTTGGTGAATAATATTAATGTAAGCGGAGATATCTGAAGCAGATGACGCTGTAGT 2940
Db 3038 TGTTTGGTGAATAATATTAATGTAAGCGGAGATATCTGAAGCAGATGACGCTGTAGT 3097
Qy 2941 GAAAAATATTTGCTCTTCTACAGAGATGTTGTTCCATATACAAATTCACCTTTTGCA 3000
Db 3098 GAAAAATATTTGCTCTTCTACAGAGATGTTGTTCCATATACAAATTCACCTTTTGCA 3157
Qy 3001 CATGACCCAGATATGTCAAAGTACAGATATTTGAACAACCTTAAAGATTTAAGATGT 3060
Db 3158 CATGACCCAGATATGTCAAAGTACAGATATTTGAACAACCTTAAAGATTTAAGATGT 3217
Qy 3061 CTTTGGTTTGTCTGGAATATTAATGCTTAAAAATGAATAACAGTACAGCTTTTATC 3120
Db 3218 CTTTGGTTTGTCTGGAATATTAATGCTTAAAAATGAATAACAGTACAGCTTTTATC 3277
Qy 3121 AGAAAGATGTGAATAATTTAAACAAACAAAGATGCCAAGACAGCATGATGACAAA 3180
Db 3278 AGAAAGATGTGAATAATTTAAACAAACAAAGATGCCAAGACAGCATGATGACAAA 3337
Qy 3181 ATGAATGAAAAAAGTACATGCTGTGATGTTGCCATGAATATCATGTCACAAAGT 3240
Db 3338 ATGAATGAAAAAAGTACATGCTGTGATGTTGCCATGAATATCATGTCACAAAGT 3397
Qy 3241 ACTACATCAGTTTGAATCTCTTAAAGACCGGACTACAGCTCGTTTCTCAGTCAA 3300
Db 3398 ACTACATCAGTTTGAATCTCTTAAAGACCGGACTACAGCTCGTTTCTCAGTCAA 3457
Qy 3301 CCTGACAGAAATTCAGTACACACCAAAATTAATCTGCTCCTGAATATGAATCTTTTC 3360
Db 3458 CCTGACAGAAATTCAGTACACACCAAAATTAATCTGCTCCTGAATATGAATCTTTTC 3517
Qy 3361 ACTCCTGAAAAACCTAAACCAACATGCTCAGAGGCTTAAACAGCACTTTCATCA 3420
Db 3518 ACTCCTGAAAAACCTAAACCAACATGCTCAGAGGCTTAAACAGCACTTTCATCA 3577
Qy 3421 GCAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAACCTTAAGCAATGCAAGACG 3480
Db 3578 GCAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAACCTTAAGCAATGCAAGACG 3637
Qy 3481 AGCTCAATTCAGACTCTCTGGAAGAAATTAAGGGGAGCTTGATATGTTCTGAATGGAT 3540
Db 3638 AGCTCAATTCAGACTCTCTGGAAGAAATTAAGGGGAGCTTGATATGTTCTGAATGGAT 3697
Qy 3541 CACAGTGAATAATGAAGTATTAACAATGCTCACCCTTGGCCGGGAAAAAAGTACAG 3600
Db 3698 CACAGTGAATAATGAAGTATTAACAATGCTCACCCTTGGCCGGGAAAAAAGTACAG 3757
Qy 3601 AGAGACGACTGTATCTTGAAGCTCTGAATTTGGAAGAGCTAGAGCAGAGAAAAAAGC 3660
Db 3758 AGAGACGACTGTATCTTGAAGCTCTGAATTTGGAAGAGCTAGAGCAGAGAAAAAAGC 3817
Qy 3661 CCCGTCACAGAACAGGAGAGAAATTAAGTATGATGATGACTTGAATGTTGTCACAGAA 3720
Db 3818 CCCGTCACAGAACAGGAGAGAAATTAAGTATGATGATGACTTGAATGTTGTCACAGAA 3877
Qy 3721 CAGAAACCTTAAAGGAGTACAGAAAGTGGAAAAAGAGCCATACGCTTCGAATCTGAT 3780
Db 3878 CAGAAACCTTAAAGGAGTACAGAAAGTGGAAAAAGAGCCATACGCTTCGAATCTGAT 3937
Qy 3781 GAACAGAGTGGCCGAGAGAAAGAGCTCAAGAAATATTAATTTGAAAAATGAAGATGAA 3840
Db 3938 GAACAGAGTGGCCGAGAGAAAGAGAGCTCAAGAAATATTAATTTGAAAAATGAAGATGAA 3997
Qy 3841 CAGAAATAGTCCGCCAAAAAAGGTTAAAGAGCCGACACCAAAACCTCTTGAGAGGT 3900
Db 3998 CAGAAATAGTCCGCCAAAAAAGGTTAAAGAGCCGACACCAAAACCTCTTGAGAGGT 4057
Qy 3901 ACACCAAAAGAAAGCCCAACATGAATGAATCTTAAAAAGAGCAAAAAAATCTGGA 3960
Db 4058 ACACCAAAAGAAAGCCCAACATGAATGAATCTTAAAAAGAGCAAAAAAATCTGGA 4117

OY	3961	CCCTCACCACGAGGGAGGAGGAAGAAAGAAAGTGGAAATACGGACACAAAG	4020
Db	4118	CTCTCCACACGACGAGGGAGGAGGAAGAAAGAAAGACAAAGTGGAAATACGGACACAAAG	4177
OY	4021	TCCAAAAGCAAAACGACGACCGAGTGTCAAGAGACACAGCAGAGAGCAATCTCTGAA	4080
Db	4178	TCCAAAAGCAAAACGACGACCGAGTGTCAAGAGAGACACAGCAGAGAGCAATCTCTGAA	4237
OY	4081	TCTAGTGCATGTGAATCTCACACAGTCCACACACACAGAAAGACAGAACCTCAAAA	4140
Db	4238	TCTAGTGCATGTGAATCTCACACAGTCCACACACACAGAAAGACAGAACCTCAAAA	4297
OY	4141	ACGCCATCACCATTCACAACCAAAAAAATGT	4172
Db	4298	ACGCCATCACCATTCACAACCAAAAAAATGT	4329
RESULT 6			
ABX14056			
ID	ABX14056	standard; cDNA: 6744 BP.	
AC			
XX	ABX14056;		
DT	24-FEB-2003	(first entry)	
XX			
DE	cDNA encoding human SCC-112 tumour suppressor gene.		
XX			
KW	ss; gene: human; chromosome 4p14; SCC-112; cancer; apoptosis; allodynia;		
KW	degenerative disorder; metastasis inhibition; breast cancer; causalgia;		
KW	kidney cancer; bladder cancer; pancreatic cancer; colon cancer;		
KW	squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;		
KW	global and focal ischemic and haemorrhagic stroke; epilepsy; neuralgia;		
KW	hypoxia-induced nerve cell damage; anxiety; diabetes mellitus;		
KW	cardiac arrest; spinal cord lesion; stomach cancer; lung cancer;		
KW	neonatal distress; Alzheimer's disease; uterine cancer; colon cancer;		
KW	multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;		
KW	Huntington's disease; Parkinson's disease; Korsakoff's syndrome;		
KW	amyotrophic lateral sclerosis; cell survival; cell proliferation;		
KW	tumour suppressor.		
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	CDS	232..4125	
FT		/*tag= a	
FT	polyA_signal	/product= "Human SCC-112"	
FT		6721..6726	
FT		/*tag= b	
FT		/standard_name= "PolyA signal"	
PN	W0200281641-A2.		
XX			
XX	17-OCT-2002.		
XX			
XX	08-APR-2002; 2002MO-US10850.		
XX			
XX	06-APR-2001; 2001US-281780P.		
PA	(GEOU) UNIV GEORGETOWN.		
XX			
XX	Kasid UN, Kumar D, Ahmad I;		
PI	WPI: 2003-103330/09.		
DR	P-PSDB: ABG72803.		
XX			
PT	New isolated SSC (undefined) tumor suppressor polypeptides and		
PT	polynucleotides, useful for diagnosing, preventing or treating cancer		
PT	or degenerative disease, e.g. Alzheimer's Disease, Huntington's		
XX	disease, or multiple sclerosis		
XX			
PS	Claim 1; Fig 1B; 83pp; English.		
CC	The invention relates to a new isolated polypeptide SCC-112. The SCC-112		

CC		polypeptides and polynucleotides are useful for diagnosing, preventing or
CC		treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
CC		pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
CC		cancer, stomach cancer, colon cancer, lung cancer); or degenerative
CC		disease or disorder (e.g. global and focal ischemic and haemorrhagic
CC		stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
CC		damage, nerve cell damage caused by cardiac arrest or neonatal distress,
CC		epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
CC		pain, causalgia, neuralgia, herpes zoster, spinal cord lesions,
CC		hyperalgesia, allodynia, Alzheimer's disease, Huntington's disease,
CC		Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
CC		Korsakoff's syndrome). The polypeptides and polynucleotides are also
CC		useful for inducing apoptosis in cancer cells, increasing survival or
CC		proliferation of a cell, or inhibiting cancer cell proliferation and/or
CC		metastasis in a cancer patient. The polynucleotides can be used as probes
CC		to detect complementary nucleotide sequences, or as primers to obtain
CC		additional copies of the polynucleotides. SSC-112 may also be used for
CC		identifying drugs for treatment of cancers. The present sequence
CC		represents cDNA encoding the human SCC-112 tumour suppressor gene which
CC		is located on chromosome 4p14.
CC		
XX		
SQ	Sequence 6744 BP; 2156 A; 1243 C; 1400 G; 1945 T; 0 other:	
	Query Match	43.5%; Score 1915.6; DB 25; Length 6744;
	Best Local Similarity	69.7%; Pred. No. 0;
	Matches 2457; Conservative	0; Mismatches 1069; Indels 0; Gaps 0;
OY	32 GAATAATTCATCCTCCGCGGGGACAGAAATTCAGATTAAATATCTAAGAAGCA	91
Db	173 GGAAGATCGCTTACCCTCCGGGGGTAAGAAGATCACGCACAATCACACGACGAGA	232
OY	92 TGGTGAGACGATTAAAGATGGTTTGGAACACTTTATGATATGAGACCAGACCTCGAA	151
Db	233 TGATCAAACGCCCTGAAGATGGTAGGAACACTTATGATATGATCAGACGATCAGAAG	292
OY	152 AAGAAAAGACACTTATTTAAACCTAGCTTACATCTTGCTTCAGATTTTTTCTCAAGC	211
Db	293 ATGAAAAACACAGATATCTCCACCTACGCTTGACCTTGACATCGAAATCTTCTCCAGGA	352
OY	212 ATCTGTGTAAGATGTTGCTTACTGATGAGACCTGCTGCTCGATATTTTTCACAGATT	271
Db	353 ACCCAATTAAGATGTGGCTGCTCCGTGATCATGTGTTGGGTGATATCTTGTGATCT	412
OY	272 ATGCTCTGAAGCTTCCTTACACATCCCTGATTAACCTAAGGATATATTTATGTTATAA	331
Db	413 ATGCCCCGGAAGCTCCATATCTACTCCCATGTGATTAACCTAAGACATATATTTTGTATATA	472
OY	332 CAAGACAGTTCAAGGGGCTAGAGATTAAGAAAGGCCCAATTCOATAGTATTTTAT	391
Db	473 CCAGCATTTAAAAGGTTTGGAGGATCAAAAGAGTCCACAGTTTAATATGATATCTTTAT	532
OY	392 TACTTGAGAACATTGCTTGGGTGCAAGTCATATAACATATGCTTGGATTAGAAAGATACA	451
Db	533 TATTAGAGAAATTTAGCTTGGGTAAATCATTAACATCTGCTTGAATTTGGAAGATGTCA	592
OY	452 ATGAATTTTACCCAGCTATACGAACACTTATTTTCAGTTATTAACAAATGGCCCAATC	511
Db	593 ATGAAATTTTATTCAGCTTTTATGAACTCTCTCCAGATCAACAAATATGCCCCAATTA	652
OY	512 AGAAAGTCATATCCACATAGTGTAGACCTTATGAGCTATATTTTGTAAAGCTGATPACG	571
Db	653 AGAAGTCAATATGCACATCTGATAGATTGATGATGATCATCATCTGAAGSTGATGAGG	712
OY	572 TGCTCAGAGACTTTTGGATACGGTTTATGTAATCTGTGACCTGCTCATAGAATTTAA	631
Db	713 TTACTCAAGAAATTTATGGGCTCCATTTCTTATTAACCTCATCTTCGACATAAAGACTTAA	772
OY	632 ACAAGCAAGCATATGATTTGGCAAAGGCTTTACGTAAGACAGACACCTCAAGCTATTTGAGC	691
Db	773 ATTAACAGACTCTTTGACCTTCAAAAAGTGTCTATTGTAAGAAAGACAGTCCAGACTATTTGAGG	832
OY	692 CATATATTACACTTTTATTAATCAGGTTCTGATGCTTGGAAAAACATCTATCAGCGATT	751

D8 833 CATGATTGCCAATTTTTCATCAATCAAGTCCTGCTGGGAAAGATCAATCAGTATG 892
QY 752 TGTGAGACAGATGCTTTGACTTAATTTTGAGCTCTACAAATATGTATGATCTTTGCTGC 811
DB 893 TGTGAAACATGATTTGATCTGATTCAGAACTTTTGTATATATCTCTCTATTAAT 952
QY 812 TCTCTGTTTACCAGCTTGATTTAAATTAAGAGCAATGATTAAGAGAGCCCTTAC 871
DB 953 TATCCGTATGCAAGCTTGATTTCAAACTTAAGAGCAATGATGAAGAGAGCATTAG 1012
QY 872 AAGTTGTAACCTACTGCAAAAATGTTGGGGCAAGAGATTCACAATGGCTCTGCAA 931
DB 1013 CTGTGTTGCACTTACTAGTAATTTGTTGGCTCCAAAGATTTCTATTTGGCAACAG 1072
QY 932 ACAAGCCACTTTGGCAGTCTACTTGGCAGAGTTTAATGATATCCATGTACCAATCCCC 991
DB 1073 ATGCGCTCTTTGGCAATGTTTCTTTGGACGATTTAATGATATTCATGTCTCTGAGAT 1132
QY 992 TGGAAATGTGAAATTTGCTAGCCATTTCTCATGACCAATCCCTGATTTAGCAAAAGCT 1051
DB 1133 TAGAAAGTGTGAAATTTGCCAGTCAATGTTTATGATCAATCCAGATTTAGGAGAGATC 1192
QY 1052 TAACAGATATCTTAAAGTAGAGTGACATGACCAAGAGAGCTATTAGACATGATGTTA 1111
DB 1193 TCACAGATATTTTAAAGTTTAAATGATCAATGATCCAGAAAGAGCTATTTGTCATGATGCA 1252
QY 1112 TTGTCTCAATAGTACAGCTGCTAAAGAGATATTTCTTGTGTCATGATCACTACTTA 1171
DB 1253 TTGTACTATATATAACGCTGCAAGAGAGAGAGCTGGCTTATGTAATGATCAGCTGCTTG 1312
QY 1172 ATTTTGTAGAGAGAGAACTTTAGCAAAAGCATGAGAGTACGCAAGAGAGCATATG 1231
DB 1313 GCTTTGTAGAGAGAAACACATGATTAACGGTGGCAGATGAAGAGAGATGATGAG 1372
QY 1232 GACTTTGCCCAATTTATAGAAATGCTTTACAGTCAAGCAGCTGGAAGAGATGCTGAA 1291
DB 1373 GTCTGCTCAGCTTTATAGAAATGCTTTCTTCAATGATGAGAGAGAGAGAGAGCTGAG 1432
QY 1292 AACAGATAGCATGATCAAGAAACAAATTCATATATATATCAAAATAGTATGTATG 1351
DB 1433 AGAAAGTACGCTGGATTAAGAGCAAACTTCTGCAATTTATATAGAAAGCATTTAG 1492
QY 1352 ATCGACTACTTGTGAAGAGATCTTGTCTCAATACATGCTTCTCAATTTAGAACTA 1411
DB 1493 ACAAACTGTGTAGAGAAATCTTGTCTCAATCTTGTCTCCCAACACCTGGAGAAAG 1552
QY 1412 CAGAGGATGAATGCTTATATTTACTTGTATGCGACACTGAGATTTAAATGCTGGAAG 1471
DB 1553 AAGAGAGATGAATGCTTATATTTACTTATATGCTAGTTTGGATCCAAATGCTGTAAG 1612
QY 1472 CATGGAATGAATGTGAATGTCAAAATCTGCTCCGACATCAAGTAAGATTTGCTTG 1531
DB 1613 CTCTCAAGAAATGGAAGTGTGCAACATGCTTCGGAAGCATATGCGAGCAATTTGG 1672
QY 1532 ACTTGATTAAGCAACCCAAACAGATGCCAGTGTGAAGCCATTTTCAAAAGTATG 1591
DB 1673 ATTTGACACAGCAGCTTACATCAGAGGCTTACTGTTCCATGCTTTGGAAGAGATG 1732
QY 1592 TTATTAAGAAATTTACTGATCTGTAAGAGCTCAGAGATTTCAATGAAGAAATTCAC 1651
DB 1733 CCATAGCAAAATTTGCTGACCCCGGAGAAAGCAAGATTTTGTGAAGAAATTTAAC 1792
QY 1652 AGGTGTTAGAGATGATGAGAAATTAAGAAAGCAGTTAGAGTACTTTAGTCCACAT 1711
DB 1793 AGTTCTCGGCGATGAGAAACTTGGTCTCAATGTGAGATTATTAATTTAGCCAACT 1852
QY 1712 GCTCTGTGAAGCAGGCTAAGGTTGTGCTGAAATTAAGAAAGTTGGGCAACCCCA 1771
DB 1853 GTTCTTGCAACAGCAATATTTGTGTGAGAGAAATGAGCCGGAACCTTCAAAATCTTA 1912
QY 1772 AACAGCTTACAAATCTTCTGGAATGATCAAGTTCTTGGAGAGAGATAGACCTG 1831
DB 1913 AGCAACCAAAATCTTTCTAGAGATGTCAAATTTCTGTGGAAGAAATCGACCTG 1972

QY 1832 TGCACATAGATACCAATCTATACAGTCTTATTTAAACAAGTGAACAAATCAATATG 1891
DB 1973 TGCACATGATTTACAGAACCAATAGTGCAGTATGTAATTTGATTAAGTCAATATGAGG 2032
QY 1892 GACAGCAGATGATTAAGATGAGGTTTCCAACTGATCAAGCCATCAGAGAGGCTTG 1951
DB 2033 GGCAGCAGATGATTAAGAGAGGAGGTTGATGTCAGATACAGCTATCCGTTCCAGAGCTTG 2092
QY 1952 AACGCTTAAGTACTTATACATATCCATGCAATTCATTTCTGCTGGAACATTTG 2011
DB 2093 AACTCTTAAGTCTTCTTTTACACATCTTACCTGCTGCTTCCACTCTGAGAGATATG 2152
QY 2012 AATCATTTACTGCTTGTGAAATGATGATGAAGAAATAGCAGAGAGCTGACATCAAA 2071
DB 2153 AGTCTTGTATACAGTCCATAGAAATGAGAGATGACAAAGGTACACAGAGCTGATTTCAA 2212
QY 2072 TTTTCAAAAACAGAGAAATTTGAAGAGATTTTTCACACATCAATGATCAGCTTGC 2131
DB 2213 TTTTGAAGAAATACAGCTCACAAATTAAGAAACAGACCTTCCAGATGATGATCAGCTTA 2272
QY 2132 TTCCGTTTTCATCACAAATCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2191
DB 2273 TTCCATTTTACATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2332
QY 2192 ATTTATTCATGCGATATTTTCTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2251
DB 2333 ACTGATATACAGCCATATTTCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2392
QY 2252 TGCATTAAGAGCTTACATCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2311
DB 2393 TCAGTAAAGAGCTGATGATCTGATGTCAGAGAACTTATTAATCTATGATTTGATG 2452
QY 2312 GTCATATTTGCTCTGCTGACACTGATCAATTTGCTGCTGCTGGAATCTTGGATGCTA 2371
DB 2453 GCCACATTTCTATGTTACAGCAGAGATGCTGCTTCTCCCAATGAAATCTGATAGCA 2512
QY 2372 CTTTCATTTGTAAGATCTTCTAGTAAATGATGATGCTTCCAGAGAGAAAGAGAGAGAG 2431
DB 2513 ATTTATTTGTAAGATCTTCTAGTAAATGATGATGATGATGATGATGATGATGATGAT 2572
QY 2432 TTTGGGTTCCAGATTAAGAGATCTCTGAGACATGCTCAAAATTTAGGCTATTTAA 2491
DB 2573 TGTGCTTCCAGATTAAGAGATCTTCCAGAGATCTGAGAGATGATGAGAGATGATGAG 2632
QY 2492 TGATGCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2551
DB 2633 TTCTGCTAAGGAGGCTGTTGGGTATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2692
QY 2552 TAAGATTTCTAACAACAATATTTGATGATGATGATGATGATGATGATGATGATGATG 2611
DB 2693 TTCTGATTTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2752
QY 2612 GTAACAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2671
DB 2753 GTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2812
QY 2672 AAGAACCTTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2731
DB 2813 AGGAACCTTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2872
QY 2732 CCCGTTACGCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2851
DB 2933 TGAAGTACTCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2992
QY 2852 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2911
DB 2993 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3052

Db 414 TGTTCGCGATGTTGGAAAACTGATGACCATAGCAAAAGAAATTTGCGACCCCGGAAA 473
QY 1624 GCTCAGAGATTTCATGAAATTTCAACAGAGTGTTAGAAGATGAGAAAATAGAAAG 1683
Db 474 GCACAAAGTTTTGTGAGAAATTTTACCAGAGTTCTCGCGGATGATGAGAACTTCGGTCT 533
QY 1684 CAGTTAGAGATCTGTGTAGTCCAAACAGTCTCTGCAAGCAGCGCTGAAGGTTGTGTGCT 1743
Db 534 CAGTTGAGATTATTAATGACCCCAACCTGTTCTTGCAAAACAAGCATATTTGTGTGAGA 593
QY 1744 GAATTAATAGAAAGTTGGGCAACCCCAAGCCTCAAAATCTTCTTCTGGAATGATC 1803
Db 594 GAAATAGCCCGGAAACTTGCAAATCTTAAGCAACCAACAAATCTTCTGAGATGGTCT 653
QY 1804 AAGTTCTCTGGAGAGATAGCAACCTGTGCACATAGATACGAAATCTATGCATCTCTT 1863
Db 654 AAATTTCTGTGGAAAGATGCGACCTGTGACATTTGATTCAGAAAGCATTAAGTGCCTA 713
QY 1864 ATTAACAAAGTAAACAAATCAATAGATGAACAGACAGATGATGAAGATGAGGTTTCCA 1923
Db 714 GTGAATTTGATGAATTAAGTCAATAGAGGGGACACAGATGATGAAGAGAGGTGTAGT 773
QY 1924 ACTGATCAAGCATCAGACAGGTCTGTAAGTCTTAAGTACTCTCATTTACACATCCC 1983
Db 774 CCAATATACGCTATCCGTTGACAGACTTGAACTTTTAAGTCTGTCTTTTACACATCT 833
QY 1984 ATCTCATTTCACTTCTGTAAGAACTTTGAATCATTTACTGCTGTCTGTAAGATGAT 2043
Db 834 ACCTGTTCCACTCTGCGAGACACATATGAGTCTTGTACAGTGTCTTAAGATGAGAGAT 893
QY 2044 GAAAAAGTAGACAGAACTGCACTACAAATTTTCAAAAAACAGAGCAAAATTTGAAG 2103
Db 894 GACAGAGTAGCAGAGAGTGTCTATTCAAATTTTGAATTAACAGTCAACAAATTTGAACA 953
QY 2104 GATTTTCCACATCAGATCAGCTGCTGCTGTTTACATCAAAATCTAAAAAGAA 2163
Db 954 GACCTTCCCGATACGATTCGACCTTAATTTCCATTTTACATCAAAAAGCAAAAGGGGT 1013
QY 2164 CCCCCCGTCAAGCAATATGCAATTCATTTGATTCATTCATGCGATATTTTCTAATAAG 2223
Db 1014 ACTCCACACCAAGCAAAACAGAGTGTGCACTGTATACAGCCATATTCACAAATTAAGAA 1073
QY 2224 ACCCAGTTTGACAGATTTTGAGCCTCTGCATTAAGAGCCTAGATCCAAAGCACTGGA 2283
Db 1074 GTCCAGTTTGACAGATTTTGAAGCTCAGTAGGATCTGAATGTGATGTGCCAATA 1133
QY 2284 CATCTCATTAACACCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2343
Db 1134 CAATTTATTAATCTCATTAGTTTCAATTGAGCCACATTTCTATGTTAGCACCAAGTCAATT 1193
QY 2344 GCTGCTCTTGGAAATCTTGGGTAGTCACTTTCAATTGGAAGATCTTCTCATAGAT 2403
Db 1194 GCTTCCCAATGAATCTGTAGTCAAAATTTTATGGAAGATCTCTAATGATGATC 1253
QY 2404 CGGTTCCAGGAAAGAGACAACTAACTTTGGGTTCCAGATGAAGAGTATCTCTGAG 2463
Db 1254 AGGTCAACAGGTGAAGAAATGAAACTGTGTCTCAGATGAAGAGTCTTCCCTCA 1313
QY 2464 ACAATGGTCAAAATTCAGGCTATTAATGAATGATGATGATGATGATGATGATGATGAT 2523
Db 1314 GTACTACCAAAAGTACAGGCAATTAATCTTGTGAAGTGTGCTGTGTGTGTGTGTGTGT 1373
QY 2524 AATCAAGTAATCAGAACTTCTACCTTAAGATGCTTAACAAATATTCATAGTATGAT 2583
Db 1374 AACCAAGTAAATCTGCAATTTCAACCTTGTTATATATCAAGGATGTGTGTATGAG 1433
QY 2584 GAGACCTTGACAGAAAGGGGAAATTAATTAACCAATATGATCACTGTGAGCTTCT 2643
Db 1434 GGTGACTGACAGAGCAAAAGATCAATTAATCTGATGATGATGATGATGATGATGAT 1493
QY 2644 GCTGGAGTGTATTTGAAGCTGCAACAAAGCACTGTTCCATGAATATCATCATTA 2703

Db 1494 GCTGCTAGTGCATTAATGAAGCTTGCTCAGGAACCTTGTACCATGAATTAATACCCA 1553
QY 2704 GAACAAATATCAGCTATGTGATTAATCAACATGATGATGATGATGATGATGATGATGAT 2763
Db 1554 GAACAGTTTCAGCTGTGACACTTGTATTAATGATGATGATGATGATGATGATGATGAT 1613
QY 2764 TTTGCCGAAACCTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAATATATGCA 2823
Db 1614 TTTGCTCAGAAAGCTCATTAAGCACTTGTGAAGTTACTGCTCCATTTGAGATATATGCG 1673
QY 2824 ATCTGTGCTTGTGCAAAAGATCTCTTAAGGAGAAAGATCAATGATGATGATGATGAT 2883
Db 1674 ATCTTGTGCTTGTGCAAAAGATCTCTTAAGGAGAAAGATCAATGATGATGATGATGAT 1733
QY 2884 TTTGGAATAATATTAATGATTAAGGCGGAGATATGCAAGCATGATGATGATGATGAT 2943
Db 1734 TTACTGAAATAATATCAGATATACGCGGAAATATCAATTAAGCAAAATCTTATGCTATG 1793
QY 2944 AAATTAATGCTCTTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3003
Db 1794 AAATTAATGCTCTTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1853
QY 3004 GACCCAGATTAATGCAAGATACAGATTTGAACAACCTTAAGATGATGATGATGATGAT 3063
Db 1854 GATCCAGATTTTACAGATCAACAGATGATGATGATGATGATGATGATGATGATGATGAT 1913
QY 3064 TGTGTTGTTCTGGAATATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3123
Db 1914 TGTGTTGTTCTGGAATATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1973
QY 3124 AAGATGTTGAATAATTAATTAACAAACAAAGATGATGATGATGATGATGATGATGATGAT 3183
Db 1974 AAGATGTTGAATAATTAATTAACAAACAAAGATGATGATGATGATGATGATGATGATGAT 2033
QY 3184 AATGAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3243
Db 2034 AATGAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2093
QY 3244 ACATACATTTGGAATCTCTTAAGACCCGCTACTACAGCTGTTCTTCTCATCACT 3303
Db 2094 TTGTCATGCAAGATTTCAACAAAGACCCAGCTGCTCCATGAATTTTATACAACT 2153
QY 3304 GACAAAGATTTCAAGTACCAACAAATTAATCTGCTCTGTAAGTAAATCTTTTCACT 3363
Db 2154 GAAAGGACTTCTGTAAGATTAAGATTAATTTTCAAGAGAAAGATGATGATGATGATGAT 2213
QY 3364 CCTGGAACCTTAACCAATGTTTGTAGAGCTGTTAACAAGCCCTTTCATCACT 3423
Db 2214 ACAGGAAAGCCAAAGCCTGCTGAGTACTAGTGCAGTAAATTAAGCCTTTATCAGCAAG 2273
QY 3424 GGCAAGCAATCTCAGACCAAAATCATCAGAAATGCAAACTGTAAGCAATGCAAGCAGC 3483
Db 2274 GGAAGGAAACCTTATGTTAGAAAGCACTGCACTGAGTGAAGCAATTAATTAATGAT 2333
QY 3484 TCAATCTCAAGCTCTCTGGAAGATTAAGGAGAGCTTGTATGATTTCTGAATATGATCAC 3543
Db 2334 TCAAGCTGGAACCTTCAACCGGAATGATCAAGGGAACAGATGATGAGAGCAGAGAA 2393
QY 3544 AGTGAATAATGAAGA 3557
Db 2394 ACTGAGTTAGTGA 2407

RESULT 8
AAA47423
ID AAA47423 standard; DNA: 3957 BP.
XX AAA47423;
AC 20-0CT-2000 (first entry)
XX
DE sequence encoding human neuron-associated protein.
XX

KM Neuron associated protein; NEUP; neurological disorder; epilepsy;
 KM ischemic cerebrovascular disease; stroke; cerebral neoplasm;
 KM Alzheimer's disease; Pick's disease; Huntington's disease;
 KM dementia; Parkinson's disease; demyelinating disease; meningitis;
 KM prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
 KM cerebral palsy; muscular dystrophy; central nervous system; CNS;
 KM peripheral nervous system; PNS; myopathy; schizophrenia;
 KM actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
 KM cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
 KM myelofibrosis; paroxysmal nocturnal hemoglobinuria; cancer;
 KM autoimmune disease; inflammation; acquired immunodeficiency syndrome;
 KM AIDS; Addison's disease; adult respiratory distress syndrome;
 KM allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
 KM Werner syndrome, trauma; human; ds.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 140..2704
 FT /tag= a
 FT /product= Neuron associated protein
 XX
 XX
 PN WO200034477-A2.
 PD 15-JUN-2000.
 XX
 XX
 PF 10-DEC-1999; 99WO-US30408.
 XX
 PR 11-DEC-1998; 98US-0210083.
 PR 11-DEC-1998; 98US-9123456.
 PR 09-FEB-1999; 99US-0119365.
 PR 16-MAR-1999; 99US-0124687.
 XX
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
 PI Lu DAM, Azimzal Y;
 XX
 DR MPI: 2000-423423/36.
 DR P-PSDB: AAB01382.
 XX
 XX
 PT New human neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders
 XX
 PS Claim 9; Page 130-131; 145pp; English.
 XX

CC complications of cancer, hemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic
 CC infections, and trauma. This sequence was given the Incyte ID no.
 CC 2888437CBI.
 XX
 SQ Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;
 Query Match 26.7%; Score 1112.8; DB 21; Length 3957;
 Best Local Similarity 68.1%; Pred. No. 3.5e-264;
 Matches 1349; Conservative 0; Mismatches 727; Indels 0; Gaps 0;
 1282 GATGCTGCAAAACAGATAGCATGATCAAAAGCAAAATGCTACATATATATCAAAAT 1341
 DB 11 111111 11 111111 111111 111111 111111 111111 11
 2 GAAGCTGCAGAAAGTACGCTGATTAAGCAAACTCTGCATATTTATATACAGAAC 61
 QY 1342 AGTATGATGATGACACTACTTGTGAAGGATCTTGTCTCAATACATGTTCCACAAAT 1401
 DB 11 111111 11 111111 111111 111111 11 11 111111
 62 ACCATTTGACGACAAACTGTTGGTAGAAGAAATCTTGTGCTACATCTTGTCCCAAC 121
 QY 1402 TTAGAAACACAGAGAGATGAATGCTATATATCTGTATGCAACACTGATTTAAAT 1461
 DB 111111 111 11111111111111111111 11111 111111 1111
 122 CTGAAACGAAAGAGAAATGTAATGCTATATATCTATGCTAGTTGATTCCAAT 181
 DB 111111 1111 11111111111111111111 1111 11 1111
 1462 GCTGTGAAAGCATTTGAATGAATGTCAAAATCTGCTCCGACATCAAGTAAAG 1521
 DB 11111 11111 11111111111111111111 11111 1111 1111
 182 GCTGTAAAGCTCTCAGCAAAATGTGGAAGTGTCAAGACATCTTCGATCCATGACGC 241
 QY 1522 GATTTGCTTGACTGATTAAGCAACCCAAACAGATGCCAGTGTCAAGGCCATATTTCA 1581
 DB 11 11111 111 11111 11111 1111 1111 1111 1111 1111
 242 GAACATATTGATTTGATTCACAAAGCAGCTCATCATCAGAGCTTAATGTTTGGCA 301
 QY 1582 AAGATGAGTATTATACAAAGAAATTTACTGATCTCGTAAGGTCAGGATTTCAATGAAG 1641
 DB 111 11111 11 1111 111111 11111 1111 11 1111 11111
 302 AAACGTGATGACCTGACAAAGAAATTTGCTGACCCCGGAAAGACAAAGATTTGGAAG 361
 QY 1642 AAATTCACACAGGTGTAGAAATGATGAGAAATTAAGAACAGTTAGAACTATTGTT 1701
 DB 11111 11111 11111 11111111111 11 11111 11 1111 11
 362 AAATTTAACCAAGTTCTCGCGATGATGAAACATTCGCTCAGTGTGAGTTATTAAT 421
 QY 1702 ACTCAACATGCTCTCGACAGCAGCTGAGGTTGTGTGCGAAATTAATCAAGATTG 1761
 DB 111111 11111 11111 1111 1111 11111 1111 1111
 422 ACCCCACCTGTTCTTGCMAACACAGATATATTTGTGAGGAATAATCCCGAAACT 481
 QY 1762 GGCAACCCCAACAGCTTACAAATCTCTCGGAATGATCAAGTTCTCTGGAGAG 1821
 DB 11 11111 11 1111111111 1111111111 1111 11111 11111
 482 GCAAATCTTACCAACCAACCAATCTTTTCTAGAGATGCTAAATTTCTGTTGAAAGA 541
 QY 1822 ATAGCACTGTGCACATAGATACCGAATCTATCAGTGTCTTATTAACAACTGAACAA 1881
 DB 11 11111111111111 1111 111111 11 1111 1111 1111 11
 542 ATCGCACTGTGCACATTTGATTCAGAGCATTAAGTGCATGTGAATGTGAATAG 601
 QY 1882 TCAATAGATGACAGCAGCATGATGAAGTGAAGGTGTTCACTGATCAAGCCATCACA 1941
 DB 1111111 1111111111 1111111 111111 11 1111 1111
 602 TCAATAGAGGAGCAGCATGATGATGAAGGAGGTGAAGTCCAGATACCTATCTCGT 661
 QY 1942 CGAGGCTTGACCTGCTTAAGTACTCTCATTTACACATCCCATCTTCAATCTCTGCT 2001
 DB 1111 11111111 1111111 11 1111111111 1111 11 111111
 662 TCAGACCTTGAACCTTCTTAAGGTTGTCTTTTACACATCTCACTCTGCTCACTCTGCA 721
 QY 2002 GAAACATTTGATCATTTACTGCTGTGAAGAAATGATGAAAGATGACAGAACT 2061
 DB 111111 1111 1111 11 1111111111 11111 11111 1111111111
 722 GAGACATATGATCTCTTGTNACAGTGCCTAAGATAGAGATGACAGATACAGAGT 781
 QY 2062 GCATCAACAATTTTCAAAACACAGCAACAAATTTGAAGAGATTTTCCACATCACA 2121
 DB 11 11111111 1111 11111 1111 1111 1111 1111 1111
 782 CTAATTTCAAAATTTTGAAGATACAGTACAAATTAAGAACAGACCTCTCCAGATPAGA 841
 QY 2122 TCAGGCTTGTCTCTGTTTATATCAACAATTAAGAAAGAGCCCGCGTCAAGCAACA 2181
 DB 11 1111 1111 1111111111 1111 11 11 1111 1111 1111
 842 TCGACCTTAATTTCCATTTTACATCAAAAGCAAGAGGAGTACACACCAACAANA 901
 QY 2182 TATGCATTCATTTATTCATGCGATATTTTCTACTAAAGAGACCCAGTTTGCACAGATA 2241
 DB 11 11111111 11 11111 11 111111 1111 1111111111

Db 902 CAGGCTGTGCATGTATACAGCCATATTCACAAATAAAGAAGTCCAGCTTGACAGATT 961
 QY 2242 TTGAGCCTCTGCATTAAGAGCCCTAGATCCAGCAACCTGGACATCTCATATACACCATG 2301
 Db 962 TTGAGCCACTCATAGAGAGTCTGATATGCTGATGCCAGAACCACTTATATACCTCATTA 1021
 QY 2302 GTTACTATTTGGTCATATTTGGCTCTCTCTGACACCTGATCAATTTTGGCTCTTGGAAATCT 2361
 Db 1022 GTTTCATTTGGGCGCCACATTTCTATGTGTAGCACAGATAGTTGGCTTCCCAATGAATCT 1081
 QY 2362 TGGGTACTCTTTCATTTGTAAGATCTTCTCATGATGATGATGGCTCCAGGGAAAAAG 2421
 Db 1082 GTAGTACCAATTTTATTTGTAAGATCTGCTAATGATGACAGGTCAACGGGGAAGAAG 1141
 QY 2422 ACAACTAAACTTTGGGTCCAGATGAGAAAGATATCTCTGAGACATGGTCAAAATTCAG 2481
 Db 1142 AATGAAAGAACTGTGGTCTCCAGATGAGAGAGTTTCCCTGGAAGTACTGCAAAAGGTACG 1201
 QY 2482 GCTATTAAATGATGGTCCATGGCTACTTGGAAATGAAAAATTAATCACAGTAAATTCAG 2541
 Db 1202 GCAATTAAACTCTGTGTAAGGTGGCTGTGGGTATGAAAAACAACCACTTAAATCTGGCC 1261
 QY 2542 ACTCTACCTTAAGATGCTTAACAACAATATGATGATGGATGGAGACTTGACAGAACAG 2601
 Db 1262 AATTCAACCTTCCGCTTATTTATCAACCATGTTGGTACTGAGGCTGACCTGACAGAGCAA 1321
 QY 2602 GGGAAATTTAGTAAACAGATATGTCAAGTCTGAGACTTGGCTGGAGTGTCTATTGTG 2661
 Db 1322 AAGAGATACAGTATGATGTATGTATGTCGCTGGGATAGCTGGGATAGCCATATAG 1381
 QY 2662 AAGTGGCACAAGAACCTGTACTACATGAATAATCATCATTTAGAACAAATATACAGTATGT 2721
 Db 1382 AAGCTTCTCGAGGAACCTGTGTACATGAATAATTTATCCCGAAGAACATTTTACACTGT 1441
 QY 2722 GCATTACTATCAACAGATGATGCTATCAATGATGAGACAGATGTTGGCCCAAAATTCAC 2781
 Db 1442 GCATTGTATTAATGATGATGATGTTACCAAGTAAAGCAGATATTTGCTCACAAGCTGAT 1501
 QY 2782 AAAGGCTTTTCCCTTACAGGCTTCCACTGATATATGAGCAATCTGTGCCCTTTGTGCA 2841
 Db 1502 AAGGCACTTGTGAAGTACTCTCCCATTTGAGATATATGGAGATCTTGGCTTGTGCC 1561
 QY 2842 AAAGATCTCTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2901
 Db 1562 AAAGATCTCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
 QY 2902 GTAAAGGCGGAGATCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2961
 Db 1622 ATAGCAGAGAGATATCATTAAGCAGATCTATGGCTACTGAGAAATTTATCACTGTG 1681
 QY 2962 CCAGAGTATGTTGCTCATATATCAATTCACCTTTTGGCAGATGACCCAGATTTATGTC 3021
 Db 1682 CCTCAATATGTAGTCCATATCATATGATTCACCTGTAGCCCATGATCCAGATTTTACA 1741
 QY 3022 GTACAGGATTTGAACAACCTTAAGAGTAAAGATCTTGTGGTTGTCTGGAATA 3081
 Db 1742 TCACAAGATGTTGATCAGCTTCTGTGATATCAAGAGTCCATGAGTGTATCTGGAAT 1801
 QY 3082 TTAATGCTTAATAATGAAATTAACAGTACACCTTTTATCAAAAGATGTTGGAATAAT 3141
 Db 1802 TTAATGACAAAGATGAAAAAATATAGCCATGCTTTTATGAAGAAAGATGCGCAGAGACAT 1861
 QY 3142 AAACAAACAAAAGATGCCCAAGAGACAGATGATGCAAAATTAATGAATAAACTGTACACT 3201
 Db 1862 AAGTTAATCCAGATATGCCCACTCTCCAGATGATGATCCAGACAAATGAATAAACTGTAT 1921
 QY 3202 GTGTGTGATGTTGCCATGATATCATCATGTCAAAGAGTACTACATACAGTTTGGATCT 3261
 Db 1922 GTATGTGATGTGGCTCTGTGTATATAATAGTAAAGATGCTTGTGCAATGACAGATTC 1981
 QY 3262 CCTAAAGACCGGTACTACAGCTGCTGTTTCTTCACTCAACTGACCAAAATTTTCACTAAC 3321
 Db 1982 CCAAGGACCCAGTCTCCCAATGAATAATTTTTTACACAACTGAAAAAGACTTCTGTAA 2041

QY 3322 ACCAAAATTAATATGCTCTCCCTGAAATGAATCATTTTTCACCTCCGGAACCTTAANA 3381
 Db 2042 GATTAAGATTATATTATGAGAAGACACAAGAGTACTTCTGTTAACAGAGAAAGCCAAAC 2101
 QY 3382 ACCAATGTTTAGAGAGCTGTTTACACAGCCACTTTTATCAGACAGGCAAGCAATCTCAG 3441
 Db 2102 GCTGAGTACTAGGTGCGAGTAAATATAGCTTTTATCAGCAACGGGAAGAAACCTTAT 2161
 QY 3442 AAATCATCAGCAATGGAAGAACTGTAAAGCAATGCAAGCAGCAGCTCAATTCAGCTCTCT 3501
 Db 2162 AGAAGCACTGGCAGCTGAGACTGGAAGCAATTTATGTAATTTAGAGCTGAACCTTCA 2221
 QY 3502 GGAAGAAATTAAGGGGAGCTTGTATGTTCTGAAATGGATCAGACAGTAAATGAAGA 3557
 Db 2222 ACCGAAATGATCAAGGAAAGAGAGTTTCAGAGGACAGACAGAACTGGAGTTAGGA 2277
 RESULT 9
 AAS44727
 ID AAS44727 standard; DNA; 714 BP.
 AC AAS44727;
 XX
 DE 18-DEC-2001 (first entry)
 XX
 XX Human full-length polynucleotide sequence #152.
 KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
 KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;
 KW cancer; lymphoma; neuroblastoma; autolymphoma disorder; cell proliferation;
 KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
 KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
 KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
 KW cytoskeletal; antirheumatic; antirheumatic; vulnery; antinflammatory;
 KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
 KW neuroprotective; osteopathic; antidiabetic; antiallergic;
 KW immunostimulant; analgesic; gene therapy.
 OS Homo sapiens.
 XX
 XX WO200164834-A2.
 PN 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04926.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 PR 17-JUN-2000; 2000US-0597707.
 PR 14-JUL-2000; 2000US-0616807.
 PR 19-SEP-2000; 2000US-0664641.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
 PI Drmanac R;
 XX
 DR WPI: 2001-589862/66.
 DR P-PSDB: AAU27827.
 XX
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries
 PT prepared from various human tissues, for diagnosis, treatment of
 PT cancer, neurological, inflammatory disorders and for use in arrays for
 PT detection -
 XX
 PS Claim 1; SEQ ID NO 152; 153bp; English.
 CC Sequences AAS44727-AAS44919 represent full-length polynucleotides and
 CC coding polynucleotides encoding polypeptides of the invention. The DNA
 CC and protein sequences are useful for the treatment, diagnosis and
 CC prevention of various types of disorder in a mammalian subject such as a

CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
 CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
 CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
 CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
 CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
 CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
 CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,
 CC cell proliferation, cell differentiation, stem cell growth factor,
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells
 CC in culture to give rise to neuroepithelial cells that can be used to
 CC augment or replace cells damaged by illness, accidental damage or genetic
 CC disorders. The sequences may also be used for regeneration of bone,
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.
 CC Note: Some sequences for this patent did not form part of the printed
 CC specification, but were obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 714 BP; 293 A; 131 C; 178 G; 112 T; 0 other;

Query Match 12.0%; Score 499; DB 22; Length 714;
 Best Local Similarity 100.0%; Pred. No. 6.1e-113;
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3674 AGGAGAGAAATAGTATGATGACTTGAATAGTTGGTACAGAGAAAGAAACCTTAAG 3733
 DB 1 AGGAGAGAAATAGTATGATGATGACTTGAATAGTTGGTACAGAGAAAGAAACCTTAAG 60
 OY 3734 GCAGTCAGGAAAGTCGAAAGAGAGCCATACGCTTCAGAAATCTGTGAAGACAGAGTGGC 3793
 DB 61 GCAGTCAGGAAAGTCGAAAGAGAGCCATACGCTTCAGAAATCTGTGAAGACAGAGTGGC 120
 OY 3794 CTGAGAGAAAGAGGCTCAAGAAAGATATTTGAAATTAAGATCAACGATAGTCCGC 3853
 DB 121 CTGAGAGAAAGAGGCTCAAGAAAGATATTTGAAATTAAGATCAACGATAGTCCGC 180
 OY 3854 CAAAAAAGGGTAAAGAGAGCCAGCCAAACCTCTGGTGGAGTACACCAAAAGAG 3913
 DB 181 CAAAAAAGGGTAAAGAGAGCCAGCCAAACCTCTGGTGGAGTACACCAAAAGAG 240
 OY 3914 AGCCAAACATGAAACTTCTAATAAAGAAAGAAAGAAATCTGGACCTCCAGCAGCAG 3973
 DB 241 AGCCAAACATGAAACTTCTAATAAAGAAAGAAAGAAATCTGGACCTCCAGCAGCAG 300
 OY 3974 AGGAGAGAGAAAGAAAGAAAGAAAGAAAGTGAATATCGGAACAGAAAGTCCAAAGCAAC 4033
 DB 301 AGGAGAGAGAAAGAAAGAAAGAAAGAAAGTGAATATCGGAACAGAAAGTCCAAAGCAAC 360
 OY 4034 AGCAGCGAGTGTCAAGAGAGAGCAGCAGAGAGAGAAATCTGCTGAATCTAGTCAATTG 4093
 DB 361 AGCAGCGAGTGTCAAGAGAGAGCAGCAGAGAGAGAGAAATCTGCTGAATCTAGTCAATTG 420
 OY 4094 AATTCACACAGTTCACACACAGAAAGAGAGAGAAAGACCATCAAAAGCGCATCCAT 4153
 DB 421 AATTCACACAGTTCACACACAGAAAGAGAGAGAAAGACCATCAAAAGCGCATCCAT 480
 OY 4154 CACACCAAAAAAATATGT 4172
 DB 481 CACACCAAAAAAATATGT 499

RESULT 10
 ABR35358
 ID ABR35358 standard; cDNA; 2496 BP.
 XX ABR35358;
 AC ABR35358;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human cDNA encoding secreted protein #496.
 XX
 KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;

KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.

OS Homo sapiens.

PN WO200177288-A2.

XX 18-OCT-2001.

XX 29-MAR-2001; 2001WO-US10224.

PR 06-APR-2000; 2000US-195562P.

XX (GENM) GENETICS INST INC.

PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 PI Gulukota K, Graham JR;

DR WPI: 2002-179321/23.

PT Five hundred and ninety two polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 PT -
 PS Claim 1; Page 328-329; 372pp; English.

CC The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABR34863-ABR35454 represent polynucleotides of the invention.

SO Sequence 2496 BP; 822 A; 430 C; 491 G; 753 T; 0 other;

Query Match 11.8%; Score 493; DB 24; Length 2496;
 Best Local Similarity 65.6%; Pred. No. 3.3e-111;
 Matches 752; Conservative 0; Mismatches 390; Indels 5; Gaps 2;

OY 2387 ATCTTCTCATGATGATGCGCTTCAGGAGAAAGCAACATACTTGGTCCAGAG 2446
 DB 83 ACCTTCACACAGATTTTGTGATCAACAGGTGAAGAAAGATGAAACCTGTCTCCAGATG 142
 OY 2447 AAGAAATATCTCCTGAGACAAATGCTCAAAATTCAGGCTATTAATAATGATGTTGATGCG 2506
 DB 143 AAGAGTTTCCCTGAAAGTACAGCAAGAGTACAGCAATTAACCTTGTGTAAGGTGCG 202
 OY 2507 TACTTGGATGAAAAATTAATCAACAGTAATTCAGGAATCTTAACTTAATGCTACCA 2566
 DB 203 TGTGGGTATGAAAAACAACAGATCTTAATCTGCCAATTCACCCCTTCGTTATTAATCAG 262
 OY 2567 CAATATTCATGATGATGAGAGCTTCAGCAGAGAGAGGAGAAATTAAGTAACAGATATGT 2626
 DB 263 CGATGTTGTTAGTAGAGGTGACCTGACAGAGCAAAAGAGATCAATTAATCTGATATGT 322
 OY 2627 CACGTCAGAGACTTGCTGCTGGAGTGTATTTGTGAAGCTGGCACAAAGACCTGTAC 2686

Db 323 CTCGGTTGGATTACTGCTGCTAGTGGCCATATATGAGCTTGTCTCAGGAACTTGTATACC 382
QY 2687 ATGAATCATCTACATTAGAACAAATATACAGTATGTGCATTAGCTATCAACGATGATGCT 2746
Db 383 ATGAATATATACCCAGAACAGTTTCAGCTCTGTGCACCTGTATATATGATGAGTGT 442
QY 2747 ATCAAGTAGACAGATGTTTGGCCAGAAACCTTCACAAAGGCCCTTCCGCTTACGGCTTC 2806
Db 443 ACCAAGTAGACAGATATTTGCTCAGAGCTGCATTAAGCGACTTGTGAAGTACTGCTCC 502
QY 2807 CACTTAGATATATGGCAATCTGTGCCCTTTGTGCAAAAGATCTGTAAAGAGAGAAAG 2866
Db 503 CATTGAGATATATGGCGATCTTGTGCTGTGTGCCAAAGATCTGTGAAGAGAGAAAG 562
QY 2867 CTCATGCTTAGGCAATGTTTGGTGAATAATATATATGTAAGGCGGAGATATGGAAGCACC 2926
Db 563 CACAGCGACAGCAAAAGTTTACTGAAAAATATCAGTATCGCAGGGGAATCATTAAGCAGA 622
QY 2927 ATGACGCTGTATAGTAAAAATATATGTCCTCTCTCCAGAGATATGTTCCATATACAA 2986
Db 623 ATCCATATGCTACTAGAAATATATATCTACTTGTGCTGAATATGATTCATTCATGTA 682
QY 2987 TTCACCTTTTGGCAGCATGACCCAGATATATGTCAAAGTACAGGATATGGAACACTTAAAG 3046
Db 683 TTCACCTGCTAGCCCATGATCCAGATTTTACAAAGATCAAGATGTGTGATCGCTTCGTG 742
QY 3047 ATGTTAAAGATATGTTTGGTTTGTGTTCTGGAATATTAATGGCTTAAAAATGAATAACA 3106
Db 743 ATATCAAAAGAGCTTATGCTTATGCTTGAAGTTTATATGACAAAGATGAACAAATA 802
QY 3107 GTCACGCTTTTATCAGAAAGATGTAGAAATATTAACAAACAAAGATGCCCAAGAC 3166
Db 803 GCCATGCTCTTATGAGAAAGATGCGAAGAACATCAAGTTTAAACAGATGCCCGCTC 862
QY 3167 CAGATGATCAAAAAATGAATGAAAACTGTACACTGTGTGATGTGTCATGAATATCA 3226
Db 863 CAGATGATCAAGCAATATGAATAAGTATACAGTATGATGTGCTCTCTGTGTTA 922
QY 3227 TCATGTCAAAGATCTCTCATATACAGTTTGAATCTCCTTAAGACCCGGTACTACGACTC 3286
Db 923 TAAATATGTAAGAGCTCTTGTGCA--ATGCAGATTTCCCAAGAGACCCGCTCAATGAA 980
QY 3287 GTTCTCTCACTCAACCTGACAGAAATTTAGTAACACAAATAATCTGCTCCTGAAA 3346
Db 981 ATTTTATAC--ACCTGAAAAGGACTCTGTGACGATTAAGGATTAATTTTCAGAGAGA 1037
QY 3347 TGAATATCAATTTTCACTCCTGAAAAACCTAAACAAACCAATGTTTACGAGCTTTAACA 3406
Db 1038 CAAGAGTACTTCTGTTAACAGSAAAGCCAAAGCCGTGAGTACTAGTGACAGTAAATA 1097
QY 3407 AGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAATTCATCAGAAATGGAACGTGTA 3466
Db 1098 AGCCTTATTCAGCAAGGGAAGAAACCTTATGTAGAAGACTGCGACTGAGATGGA 1157
QY 3467 GCAATGCAAGAGCACTCAATCCAACTCCTCTGGAAGATTAAGGGAGGCTGTGTA 3526
Db 1158 GCAATATTAATGTAATTCAGAGCTGAACCTTCAACCGGAATTCATCAAGGGAACAGA 1217
QY 3527 GTTCTGA 3533
Db 1218 GTTCAGA 1224

RESULT 11
AAK85948
ID AAK85948 standard; DNA; 7943 BP.
XX
AC AAK85948;
XX
DF 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40760.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01354.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR	21-SEP-2000	200005-0234274
PR	23-SEP-2000	200005-0234397
PR	23-SEP-2000	200005-0234998
PR	26-SEP-2000	200005-0235484
PR	27-SEP-2000	200005-0235834
PR	27-SEP-2000	200005-0235836
PR	29-SEP-2000	200005-0235627
PR	29-SEP-2000	200005-0236367
PR	29-SEP-2000	200005-0236369
PR	29-SEP-2000	200005-0236368
PR	02-OCT-2000	200005-0236602
PR	02-OCT-2000	200005-0237037
PR	02-OCT-2000	200005-0237039
PR	02-OCT-2000	200005-0237040
PR	13-OCT-2000	200005-0239353
PR	13-OCT-2000	200005-0239357
PR	20-OCT-2000	200005-0240960
PR	20-OCT-2000	200005-0241221
PR	20-OCT-2000	200005-0241785
PR	20-OCT-2000	200005-0241786
PR	20-OCT-2000	200005-0241877
PR	20-OCT-2000	200005-0241808
PR	20-OCT-2000	200005-0241806
PR	20-OCT-2000	200005-0241826
PR	01-NOV-2000	200005-0244617
PR	08-NOV-2000	200005-0246474
PR	08-NOV-2000	200005-0246475
PR	08-NOV-2000	200005-0246476
PR	08-NOV-2000	200005-0246477
PR	08-NOV-2000	200005-0246478
PR	08-NOV-2000	200005-0246523
PR	08-NOV-2000	200005-0246524
PR	08-NOV-2000	200005-0246526
PR	08-NOV-2000	200005-0246527
PR	08-NOV-2000	200005-0246528
PR	08-NOV-2000	200005-0246532
PR	08-NOV-2000	200005-0246609
PR	08-NOV-2000	200005-0246610
PR	08-NOV-2000	200005-0246611
PR	08-NOV-2000	200005-0246613
PR	17-NOV-2000	200005-0249207
PR	17-NOV-2000	200005-0249208
PR	17-NOV-2000	200005-0249209
PR	17-NOV-2000	200005-0249210
PR	17-NOV-2000	200005-0249211
PR	17-NOV-2000	200005-0249212
PR	17-NOV-2000	200005-0249213
PR	17-NOV-2000	200005-0249214
PR	17-NOV-2000	200005-0249245
PR	17-NOV-2000	200005-0249264
PR	17-NOV-2000	200005-0249265
PR	17-NOV-2000	200005-0249266
PR	17-NOV-2000	200005-0249297
PR	17-NOV-2000	200005-0249299
PR	17-NOV-2000	200005-0249300
PR	01-DEC-2000	200005-0250160
PR	01-DEC-2000	200005-0250391
PR	05-DEC-2000	200005-0251030
PR	05-DEC-2000	200005-0251988
PR	05-DEC-2000	200005-0256179
PR	08-DEC-2000	200005-0251856
PR	08-DEC-2000	200005-0251868
PR	08-DEC-2000	200005-0251889
PR	08-DEC-2000	200005-0251989
PR	08-DEC-2000	200005-0251990

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0235678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
DR WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 40760; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM821170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87654 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.

Sequence 7943 BP; 2567 A; 1175 C; 1468 G; 2733 T; 0 other:

	Query Match	10.5%	Score 438;	DB 22;	Length 7943;	
	Best Local Similarity	85.3%;	Pred. No. 2.le-97;			
	Matches 541; Conservative	0;	Mismatches 0;	Indels 93;	Gaps 1;	
OY	3633 GGAGAGCGCTTGAGGACGAGAAAAAAGCCCGTCACAGAACAGGAGGAGAATTTAGTAT					3692
Dd	1 GGAGAAAGCCTTAGAGAGCGAGAAAAAACCCTGCACAGAACAGGAGGAATTATGGTAT					60
OY	3693 GGATGACTTGACTAAGTTGGTACAGGAACAGAAAACCTAAAGCAATCAGCGAAGTCGAA					3752
Dd	61 GGATGACTTGACTAAGTTGGTACAGGAACAGAAAACCTAAAGCAATCAGCGAAGTCGAA					120
OY	3753 AAGAGGCCATPACGGCTTCAGAAATTCGATGAACACGACGTGGCTGAGGAAAAAGGCTCAA					3812
Dd	121 AAGAGGCCATPACGGCTTCAGAAATTCGATGAACACGACGTGGCTGAGGAAAAAGGCTCAA					180
OY	3813 AGAGATATATTTAAGAAAAATGAAGATGAAGACAGATAGTCCGCCAAAAAAGSGTTAAAGAGG					3872
Dd	181 AGAAGATATATTTAAGAAAAATGAAGATGAAGACAGATAGTCCGCCAAAAAAGSGTTAAAGAGG					240
OY	3873 CCGACCAACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAAGGCCAACATGAAAACTTC					3932
Dd	241 CCGACCAACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAAGGCCAACATGAAAACTTC					300
OY	3933 TAAAAAAGGAAGCAAAAAAAAAAATCTGCACTCCGACCCAGAGGAGGAGGAAGAAAGA					3992
Dd	301 TAAAAAAGGAAGCAAAAAAAAAAATCTGCACTCCGACCCAGAGGAGGAGGAAGAAAGA					360
OY	3993 AAGACAAGTGGAAATAACGGAACAGAGTCCCAAAAGCAAAACAGCACCGAGTGTCAAGAG					4052
Dd	361 AAGACAAGTGGAAATAACGGAACAGAGTCCCAAAAGCAAAACAGCACCGAGTGTCAAGAG					420
OY	4053 AGCACAGACAG-----					4062
Dd	421 AGCACAGACAGAGGTAAAGCATGTCTAACTCTAACTGATCTGTTTTGTTACTATATTATA					480
OY	4063 -----					4079

Db	481	AATCATATAATTGGATGCACTATCCACACTTTGGGTCTTCCCCAAAGCAGACAGACAATTCCTGA	540
QY	4080	ATCTAGTGCAATTAATTCACACAGTCGCACACACAGAAAAGACAGAGAACCATCAA	4139
Db	541	ATCTAGTGCAATTAATTCACACAGTCGCACACACAGAAAAGACAGAGAACCATCAA	600
QY	4140	AAGCCATNCNCAATCACACACCACAAAAAATGTG	4173
Db	601	AAGCCATNCNCAATCACACACCACAAAAAATGTG	634
 RESULT 12 AAV87629 standard; CDNA: 439 BP.			
ID	AAV87629	standard; CDNA: 439 BP.	
XX	AAV87629;		
AC			
DT	12-FEB-1999	(first entry)	
XX			
DE	EST clone DY17.		
XX			
KW	Expressed sequence tag; secreted protein; haematopoiesis regulator;		
KW	tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;		
KW	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;		
KM	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO9845437-A2.		
PD	15-OCT-1998.		
XX			
PF	10-APR-1998;	98WO-US06956.	
PR	10-APR-1997;	97US-0837312.	
XX			
PA	(GENE) GENETICS INST INC.		
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;		
PI	Racie LA, Spaulding V, Treacy M,		
XX	WPt: 1999-070078/06.		
DR			
XX	New polynucleotides encoding human secreted proteins - derived from		
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,		
PT	ovary, pituitary, rectum and colon cDNA libraries		
PS	Claim 1; Page 126; 641pp; English.		
XX			
CC	The present sequence represents an expressed sequence tag (EST), and is		
CC	a polynucleotide of the invention. The polynucleotides of the invention		
CC	are all secreted EST sequences isolated from a variety of human tissue		
CC	sources. The EST sequences and proteins encoded by them are predicted to		
CC	have useful biological activities which would make them suitable for		
CC	treating, preventing or ameliorating medical conditions in humans and		
CC	animals, although no supporting data is given. Suggested activities		
CC	include nutritional activity, immune stimulating or suppressing activity,		
CC	haematopoiesis regulating activity, tissue growth activity,		
CC	actinin/inhibin activity, chemotactic/chemokinetic activity, hemostatic		
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory		
CC	activity, cadherin/tumour invasion suppressor activity, tumour inhibition		
CC	activity. The EST sequences are also stated to be useful for gene		
CC	therapy.		
XX			
SQ	Sequence 439 BP; 150 A; 101 C; 85 G; 103 T; 0 other;		
QY	Query Match	9.9%; Score 411.4; DB 20; Length 439;	
	Best local Similarity	98.6%; Pred. No. 2.le-91;	
	Matches 415; Conservative 0; Mismatches 6; Indels 0; Gaps 0;		
Db	3216 CATGAAATTCATCATGTCAAGAAGTAGTACACATACAGTTTGGAAATCTCTAAAGACCGGT	3275	
	13 CATGGCCCTCACTCATGTCTAAAGAGTAGTACACTAAGTTTGGAAATCTCTAAAGACCGGT	72	

OY	3276	ACTACCAAGCTGTTCTTCTCACTCAACCTGACAGAAATTTCACTACACCAAAATTAATCT	3335
Db	73	ACTACCAAGCTGTTCTTCTCACTCAACCTGACAGAAATTTCACTACACCAAAATTAATCT	132
OY	3336	GCCACCGAATGAAATCATTTTTCAGTCCCTGGAAAACTTAAACCAACCAATGTTCTAGG	3395
Db	133	GCTCTCTGAAATGAAATCATTTTTCAGTCCCTGGAAAACTTAAACCAACCAATGTTCTAGG	192
OY	3396	AGCTGTTAAACAAGCACCTTTCATCAGCAGGCAAGCAATTCCTAGACCAATTCATCAGCAAT	3455
Db	193	AGCTGTTAAACAAGCACCTTTCATCAGCAGGCAAGCAATTCCTAGACCAATTCATCAGCAAT	252
OY	3456	GGAACACGTGAAGCAATGCAAGCAGCAGCTCAATTCCAAAGCTCTCTGGAAAGATAAAGG	3515
Db	253	GGAACACGTGAAGCAATGCAAGCAGCAGCTCAATTCCAAAGCTCTCTGGAAAGATAAAGG	312
OY	3516	GAGCCTTGATGTGTTCTGTAATGATGCACAGTGAAGAAATGAAGATATACCAATGTCTTCAC	3575
Db	313	GAGCCTTGATGTGTTCTGTAATGATGCACAGTGAAGAAATGAAGATATACCAATGTCTTCAC	372
OY	3576	TTTTCACCGGGGAAAAAAGTGACAAGAGAGACGACTGTGATCTTTGAAGGTCGAATTGGA	3635
Db	373	TTTTCACCGGGGAAAAAAGTGACAAGAGAGACGACTGTGATCTTTGAAGGTCGAATTGGA	432
OY	3636	G 3636	
Db	433	G 433	
RESULT 13			
ABL89669	standard; cDNA: 461 BP.		
XX	ABL89669;		
XX	24-MAY-2002 (first entry)		
DE	Human polynucleotide SEQ ID NO 231.		
XX			
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;		
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;		
KW	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;		
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; human; secreted protein; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200190304-A2.		
XX			
PD	29-NOV-2001.		
XX			
PF	18-MAY-2001; 2001WO-US16450.		
XX			
PR	19-MAY-2000; 2000US-205515P.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Birse CE, Rosen CA;		
XX			
DR	WP1; 2002-122018/16.		
XX			
DR	P-PSDB; ABB89260.		
XX			
PT	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and		
PT	prevention of neural, immune system, muscular, reproductive, pro-		
PT	gastrointestinal, pulmonary, cardiovascular, renal and proliferative		
PT	disorders -		
XX			
PS	Claim 4; SEQ ID NO 231: 2081bp + Sequence Listing: English.		
XX			
CC	The invention relates to novel genes (ABL89449-ABL90853) and proteins		
CC	(ABB89040-ABB90444) useful for preventing, treating or ameliorating		
CC	medical conditions e.g. by protein or gene therapy. The genes are		

CC Isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 461 BP; 135 A; 84 C; 116 G; 121 T; 5 other;

Query Match 8.5%; Score 353.4; DB 24; Length 461;
 Best Local Similarity 98.9%; Pred. No. 4.6e-77;
 Matches 354; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCTCATTCAGAGCTAGAGCAATGATGAGAAATTCATATCCCGCTGGGCTCAAG 60
 |||||||
 DB 104 ATGGCTCATTCAGAGCTAGAGCAATGATGAGAAATTCATATCCCGCTGGGCTCAAG 163
 |||||||
 QY 61 GAAATATCAGATTAATATCTTAAGAGAGATGCTGAGACGATTAAGATGTTGTA 120
 |||||||
 DB 164 GAAATATCAGATTAATATCTTAAGAGAGATGCTGAGACGATTAAGATGTTGTA 223
 |||||||
 QY 121 ACTTTATGATATGAGACGAGACTCTGAAGAGAAAGAGCTTTATTTAACTACGT 180
 |||||||
 DB 224 ACTTTATGATATGAGACGAGACTCTGAAGAGAAAGAGCTTTATTTAACTACGT 283
 |||||||
 QY 181 TTACATCTGCTGCATATTTTTCACAGCACTCCGTAAGATGTTGCTTACGTGTA 240
 |||||||
 DB 284 TTACATCTGCTGCATATTTTTCACAGCACTCCGTAAGATGTTGCTTACGTGTA 343
 |||||||
 QY 241 GCGTGGCTGCTGCATATTTTTCACAGATTTATGCTCGAAGCTCTTACACATCCCT 300
 |||||||
 DB 344 GCGTGGCTGCTGCATATTTTTCACAGATTTATGCTCGAAGCTCTTACACATCCCT 403
 |||||||
 QY 301 GATAAAGTAAAGATATATTTATGTTATTAACAGACAGTGAAGGGCTAGAGATA 358
 |||||||
 DB 404 GATAAAGTAAAGATATATTTATGTTATTAACAGACAGTGAAGGGCTAGAGATA 461
 |||||||

RESULT 14
 ABS56684
 ID ABS56684 standard; cDNA; 1104 BP.
 XX
 AC ABS56684;
 XX
 DT 23-JAN-2003 (first entry)
 XX
 DE Euchromosome fragile intelligence delay syndrome protein 20.02 cDNA.
 XX
 KM Euchromosome fragile intelligence delay syndrome protein 20.02; human;
 XX
 KM euchromosome fragile intelligence delay syndrome; dementia; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FT Key Location/Qualifiers
 FT CDS 184..732
 FT /tag- a
 FT /product- "euchromosome fragile intelligence delay
 syndrome-associated protein, 20.02".
 XX
 PN CN1351041-A.
 XX
 PD 29-MAY-2002.
 XX
 PF 26-OCT-2000; 2000CN-0125797.

XX
 PR 26-OCT-2000; 2000CN-0125797.
 XX
 PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
 XX
 PI Mao Y, Xie Y;
 XX
 DR WPI; 2002-637117/69.
 DR P-PSDB; ABB84560.
 PT Human euchromosome fragile intelligence delay syndrome associated
 PT protein 20.02 polypeptide, used to treat e.g. dementia -
 PS
 PS Claim 6: Page 29-30 (disclosure); 33pp: Chinese.
 XX
 CC This invention describes the novel human euchromosome fragile
 CC intelligence delay syndrome associated protein, 20.02. The polypeptide is
 CC used in treating diseases such as euchromosome fragile intelligence delay
 CC syndrome, and dementia. This sequence encodes the human euchromosome
 CC fragile intelligence delay syndrome-associated protein, 20.02 described
 CC in the disclosure of the invention.

Query Match 8.4%; Score 350.4; DB 24; Length 1104;
 Best Local Similarity 69.7%; Pred. No. 3.7e-76;
 Matches 488; Conservative 0; Mismatches 211; Indels 1; Gaps 1;

QY 1041 AGCAAAAGACTTAACAGAGTATCTTAAGAGTACATGACCTTGAGAAAGTATTA 1100
 |||||
 DB 1 AGCAAAAGACTTAACAGAGTATCTTAAGAGTACATGATGATCCAGAAAGCTATTCG 60
 |||||
 QY 1101 ACATGATGATATGCTCAATAGTATGCTCAATAGTATGCTCAATAGTATGCTCAATAG 1160
 |||||
 DB 61 TCATGATGATATGCTCAATAGTATGCTCAATAGTATGCTCAATAGTATGCTCAATAG 119
 |||||
 QY 1161 TCATGATGATATGCTCAATAGTATGCTCAATAGTATGCTCAATAGTATGCTCAATAG 1220
 |||||
 DB 120 TCATGATGATATGCTCAATAGTATGCTCAATAGTATGCTCAATAGTATGCTCAATAG 179
 |||||
 QY 1221 AGCATGATGAGGAGCTGCCAATTTATTAAGAAATATGCTTTACAGTACAGCTGGAAA 1280
 |||||
 DB 180 AGCATGATGAGGAGCTGCCAATTTATTAAGAAATATGCTTTACAGTACAGCTGGAAA 239
 |||||
 QY 1281 AGATGCTGCAAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1340
 |||||
 DB 240 GGAAGCTGCAAGAAAGTCACTGATTAAGAGACAACTTCTGATATTTATTAATGAGA 299
 |||||
 QY 1341 TAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
 |||||
 DB 300 CAGCATTTAGACGAATAATTTGTTGTAAGAAATCTTTGCTCAGATATCTTGTCCCAACA 359
 |||||
 QY 1401 TTTAGAACTACAGAGAGATGAATGCTTATATTAATCTTATGCTTATGCTTATGCTTAT 1460
 |||||
 DB 360 CCTGGAACAGAGAGAGATGAATGCTTATATTAATCTTATGCTTATGCTTATGCTTAT 419
 |||||
 QY 1461 TCGCTGCAAAAGATTAAGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 1520
 |||||
 DB 420 TCGCTGCAAAAGATTAAGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 479
 |||||
 QY 1521 GATTTGCTTGAATGATTAAGCAAAACAGATGCTGAGTGAAGGCTATGATTTTC 1580
 |||||
 DB 480 CGAATGATGATGATTAAGCAAAACAGATGCTGAGTGAAGGCTATGATTTTC 539
 |||||
 QY 1581 AAAAGTATGATGATTAAGCAAAACAGATGCTGAGTGAAGGCTATGATTTTC 1640
 |||||
 DB 540 AAAAGTATGATGATTAAGCAAAACAGATGCTGAGTGAAGGCTATGATTTTC 599
 |||||
 QY 1641 GAAATTCACAGAGCTTTAGAGATGATGAGAAATTAAGAAAGCAATTTGAAGTACTTGT 1700
 |||||
 DB 600 GAAATTTAACCAAGGCTTCTGCGGATGATGAGAAATTTGCTCAATGATGAGATTTATAT 659
 |||||
 QY 1701 TAGTCCACATGCTCTCTGCAAGCAGGCTGAAGGTTGTGTG 1740

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 08:52:32 ; Search time 14865 Seconds
(without alignments)
11484.418 Million cell updates/sec

Title: US-09-512-581B-3

Perfect score: 4173
Sequence: 1 atgctcattcaagaactag.....cacaacacaaaaaatgtg 4173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_inv:*
33: em_htg_other:*
34: em_htg_mus:*
35: em_htg_pln:*
36: em_htg_rnd:*
37: em_htg_mam:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4165.6	99.8	5309	9	AB023196	AB023196 Homo sapi
2	4165.6	99.8	7444	9	HS26H231	AL137201 Novel hum
3	4141	99.2	5253	9	HS095825	U95825 Human andro
4	3471.8	83.2	5317	10	AY102267	AY102267 Mus muscu
5	2816.6	67.5	5944	10	AK122414	AK122414 Mus muscu
6	1815.6	43.5	6744	9	AF294791	AF294791 Homo sapi
7	1754.4	42.0	1852	9	HS050533	U50533 Human BRCA2
8	1464.6	35.1	4252	9	BC039256	BC039256 Homo sapi
9	1041.4	25.0	5177	9	AB014548	AB014548 Homo sapi
10	937.6	22.5	1976	10	BC021408	BC021408 Mus muscu
11	932.2	22.3	2265	9	AK098331	AK098331 Homo sapi
12	932.2	22.3	2705	9	BC041361	BC041361 Homo sapi
13	447	10.7	2079	9	AK026889	AK026889 Homo sapi
14	447	10.7	137246	9	HS49010	284572 Human DNA S
15	447	10.7	168487	9	AL138820	AL138820 Human DNA
16	443	10.6	72157	2	AC016449	AC016449 Homo sapi
17	411.4	9.9	439	6	BD059747	BD059747 Secretd
18	389.8	9.3	54398	2	AC068224	AC068224 Homo sapi
19	332.4	8.0	236508	2	AC111126	AC111126 Mus muscu
20	292	7.0	295	6	BD026562	BD026562 Sequence
21	288.6	6.9	263479	2	AC125986	AC125986 Rattus no
22	288	6.9	3779	9	BC009650	BC009650 Homo sapi
23	232.2	5.6	772	6	BD159124	BD159124 Primer fo
24	232.2	5.6	2212	9	AK021757	AK021757 Homo sapi
25	232.2	5.6	227	6	AX494307	AX494307 Sequence
26	208.8	5.0	54398	2	AC068224	AC068224 Homo sapi
27	207.4	5.0	113704	9	HS267P19	Z75889 Human DNA s
28	206.8	5.0	120652	9	AL353724	AL353724 Human DNA
29	206.8	4.3	85592	2	AL512630	AL512630 Mus muscu
30	179.6	4.3	179132	10	AL358892	AL358892 Mouse DNA
31	179.6	4.3	236508	2	AC111126	AC111126 Mus muscu
32	179.6	4.2	110000	2	AC094469_0	AC094469 Rattus no
33	173.2	3.4	184321	2	AC110493	AC110493 Mus muscu
34	142.8	3.4	190058	2	AC112263	AC112263 Mus muscu
35	142.8	3.4	298053	2	AC116574	AC116574 Mus muscu
36	142.8	3.4	233589	2	AC095342	AC095342 Rattus no
37	142.6	3.1	98856	9	AC113150	AC113150 Homo sapi
38	131	3.1	145450	2	AC068352	AC068352 Homo sapi
39	131	3.1	160962	2	AC022862	AC022862 Homo sapi
40	131	3.1	167671	2	AC090081	AC090081 Homo sapi
41	131	3.0	69740	2	AC087748	AC087748 Homo sapi
42	124	2.9	145450	2	AC068352	AC068352 Homo sapi
43	121.4	2.9	162775	2	AC023006	AC023006 Homo sapi
44	121.4	2.9	173613	3	AC007475	AC007475 Drosophila
45	119.2	2.9	173613	3	AC007475	AC007475 Drosophila

ALIGNMENTS

RESULT 1
AB023196 5309 bp mRNA linear PRI 10-MAY-2002
LOCUS AB023196
DEFINITION Homo sapiens mRNA for KIAA0979 protein, partial cds.
ACCESSION AB023196
VERSION AB023196.2 GI:20521717
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.

XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
 JOURNAL DNA Res. 6 (1), 63-70 (1999)
 MEDLINE 99246063
 PUBMED 10231032
 REFERENCE 2 (bases 1 to 5309)
 AUTHORS Ohara, O., Nagase, T. and Kikuno, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: cdna@info.kazusa.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)
 On May 9, 2002 this sequence version replaced gi:4589601.

COMMENT
 FEATURES
 source

1..5309
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="hj07056s1"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pbluescriptII SK plus"
 /dev_stage="adult"
 /note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 722 was derived from fj09522 and 723 - 5309 was derived from hj07056)."
 1..5309
 /gene="KIAA0979"
 <1..4452
 /gene="KIAA0979"
 /codon_start=1
 /product="KIAA0979 protein"
 /protein_id="BAA76823.2"
 /db_xref="GI:20521718"

gene
 CDS

BASE COUNT
 ORIGIN

Query Match 99.8%; Score 4165.6; DB 9; Length 5309;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCTATTCGAAGACTAGAGCAATGATGGAATAATACATATCCGCTGGGTCAAG 60
 Db 109 ATGGCTATTCGAAGACTAGAGCAATGATGGAATAATACATATCCGCTGGGTCAAG 168
 QY 61 GAAATATCAGATAATAATATCTAAAGAGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120

Db 169 GAAATATCAGATAATAATATCTAAAGAGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 228
 QY 121 ACTTTATGATATGAGCCAGAGACTGTGAGAAAGAAAGAGCTTTTAAACCTAGCT 180
 Db 229 ACTTTATGATATGAGCCAGAGACTGTGAGAAAGAAAGAGCTTTTAAACCTAGCT 288
 QY 181 TTACATCTTGCTGATGATATTTTCTCAAGCATCTGTGAAGATGTTCCCTACTGTA 240
 Db 289 TTACATCTTGCTGATGATATTTTCTCAAGCATCTGTGAAGATGTTCCCTACTGTA 348
 QY 241 GCGTGTGCTTGTGATATTTTCAGAGATTTATGCTCTGAGCTCCTTACACATCCCT 300
 Db 349 GCGTGTGCTTGTGATATTTTCAGAGATTTATGCTCTGAGCTCCTTACACATCCCT 408
 QY 301 GATAACTGAAGGATATATTTATGTTTATACAGACAGTGAAGGGGCTGAGAGATACA 360
 Db 409 GATAACTGAAGGATATATTTATGTTTATACAGACAGTGAAGGGGCTGAGAGATACA 468
 QY 361 AAGAGCCCAATTCATATAGTATTTTATTTACTTGAGAACATTTGCTGGGTCAAGTCA 420
 Db 469 AAGAGCCCAATTCATATAGTATTTTATTTACTTGAGAACATTTGCTGGGTCAAGTCA 528
 QY 421 TATACATATGCTTTGACTTGAAGATATGCAATGAATTTTACCCAGCTATACAGAAC 480
 Db 529 TATAACATATGCTTTGAGTTGAAGATGCAATGAATTTTACCCAGCTATACAGAAC 588
 QY 481 TTATTTTACGTTATTAACCAATGGCCACCAATCAGAAAGCCATGACATGTTGAGCTT 540
 Db 589 TTATTTTACGTTATTAACCAATGGCCACCAATGCAATGCAATGTTGAGCTT 648
 QY 541 ATGAGCTCTATTTATTTGTAAGGTGATATACAGTGTCTCAGAGCTTTTGATACGTTTGA 600
 Db 649 ATGAGCTCTATTTATTTGTAAGGTGATATACAGTGTCTCAGAGCTTTTGATACGTTTGA 708
 QY 601 GTAATCTGGTACTGCTCATTAAGATTTTAAACAAGACAGATATGATTTGGCAAAGCT 660
 Db 709 GTAATCTGGTACTGCTCATTAAGATTTTAAACAAGACAGATATGATTTGGCAAAGCT 768
 QY 661 TTACTGAAGAGACAGCTCAAGTATTTAGAGCCATATATATCCATTTTAAACAGATT 720
 Db 769 TTACTGAAGAGACAGCTCAAGTATTTAGAGCCATATATATTTAAACAGATT 828
 QY 721 CTGATGCTTGGGAAACATCTATCAGCGATTGTGACAGCATGCTTTGACTTAATTTTG 780
 Db 829 CTGATGCTTGGGAAACATCTATCAGCGATTGTGACAGCATGCTTTGACTTAATTTTG 888
 QY 781 GAGCTCTACATATTTGATATGATATTTGCTGCTCTGTTTATCCCACTTGAATTTTAA 840
 Db 889 GAGCTCTACATATTTGATATGATATTTGCTGCTCTGTTTATCCCACTTGAATTTTAA 948
 QY 841 TTAAGAGCATGATTAATGAGAGCGGCTTACAGTGTGTTAACTACAGGCAAAATGTTT 900
 Db 949 TTAAGAGCATGATTAATGAGAGCGGCTTACAGTGTGTTAACTACAGTGTGCAAAATGTTT 1008
 QY 901 GGGGCAAAAGATTCAGATTTGGCTTCTCAAAACAAGCCACTTTGGCAGTGTACTTGGGC 960
 Db 1009 GGGGCAAAAGATTCAGATTTGGCTTCTCAAAACAAGCCACTTTGGCAGTGTACTTGGGC 1068
 QY 961 AGGTTTAAATGATATTCATATGATACCAATCCGCTGGAATGTGGAATTTTGTACCAATTT 1020
 Db 1069 AGGTTTAAATGATATTCATATGATACCAATCCGCTGGAATGTGGAATTTTGTACCAATTT 1128
 QY 1021 CTCATGAACCATTCGATTTTACGAAAGACTTAACAGATGTTCTTAAAGTAGGTCACAT 1080
 Db 1129 CTCATGAACCATTCGATTTTACGAAAGACTTAACAGATGTTCTTAAAGTAGGTCACAT 1188
 QY 1081 GACCCCTGAGAGAGTATTTAGACATGATGTTATTTGTCAATATGTTTACAGCTGCTTAAAG 1140
 Db 1189 GACCCCTGAGAGAGTATTTAGACATGATGTTATTTGTCAATATGTTTACAGCTGCTTAAAG 1248
 QY 1141 GATATTTCTTGTGCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTTAGACAAA 1200

Db 1249 GATATTCTTGCTGTCATGATCATCTTAAATTTTGTGAGAGAGAAACATTAGACAAA 1308
QY 1201 CGATGAGAGTACGCAAGAGCCATGATGGGACTGGCCAAATTTATAGAATATGCT 1260
Db 1309 CGATGAGAGTACGCAAGAGCCATGATGGGACTGGCCAAATTTATAGAATATGCT 1368
QY 1261 TTACAGTCCAGAGCTGGAAGAGTGTGCAAAAACAGATAGCATGATCAAAAGCAATTTG 1320
Db 1369 TTACAGTCCAGAGCTGGAAGAGTGTGCAAAAACAGATAGCATGATCAAAAGCAATTTG 1428
QY 1321 CTACATATATATATCAAAAAATAGTATGTGATGCACTGTTGTAAGCGATCTTGGCT 1380
Db 1429 CTACATATATATATCAAAAAATAGTATGTGATGCACTGTTGTAAGCGATCTTGGCT 1488
QY 1381 CAATACATGTTCTTCACATTTAGAAACTACAGAAACGATGGAATGCTTATATTTCTTG 1440
Db 1489 CAATACATGTTCTTCACATTTAGAAACTACAGAAACGATGGAATGCTTATATTTCTTG 1548
QY 1441 TATGCCACACTGGATTTAAATGCTGGAAGCACTGTAATGAAATGTGAAATGTCAAAAT 1500
Db 1549 TATGCCACACTGGATTTAAATGCTGGAAGCACTTAAATGAAATGTGGAATGTCAAAAT 1608
QY 1501 CTGCTCCGACATCAAGTAAAGATTTGCTTGACTTGATTAAGCAACCCAAACAGATGCC 1560
Db 1609 CTGCTCCGACATCAAGTAAAGATTTGCTTGACTTGATTAAGCAACCCAAACAGATGCC 1668
QY 1561 AGTGTCAAGCCATATTTCAAAAGTGATGGTTATTACAAGAAATTTACTGATCTCTGGT 1620
Db 1669 AGTGTCAAGCCATATTTCAAAAGTGATGGTTATTACAAGAAATTTACTGATCTCTGGT 1728
QY 1621 AAGGCTCAGAGATTTTCATGAGAAATTCACACAGGTTGTAGAGATGATGAGAAATTAAGA 1680
Db 1729 AAGGCTCAGAGATTTTCATGAGAAATTCACACAGGTTGTAGAGATGATGAGAAATTAAGA 1788
QY 1681 AAGCAATTGAAAGTACTTGTAGTCCACATGCTCCTGCAAGCAGCTGAAGGTGTGTC 1740
Db 1789 AAGCAATTGAAAGTACTTGTAGTCCACATGCTCCTGCAAGCAGCTGAAGGTGTGTC 1848
QY 1741 CGTGAATTAACATAGAGTTGGGCAACCCCAACAGCCTACAAATCCCTTCGAGAAAG 1800
Db 1849 CGTGAATTAACATAGAGTTGGGCAACCCCAACAGCCTACAAATCCCTTCGAGAAAG 1908
QY 1801 ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCACATAGATACCAGATCTATCACTGCT 1860
Db 1909 ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCACATAGATACCAGATCTATCACTGCT 1968
QY 1861 CTTATTAAACAAGTGAACAATCATATAGTGAACAGCAGATGATGAAGATGAGGTGTT 1920
Db 1969 CTTATTAAACAAGTGAACAATCATATAGTGAACAGCAGATGATGAAGATGAGGTGTT 2028
QY 1921 CCAACTGATCAAGCATCGAGCAGGCTTGAACTGCTTAGCTACTTCATTTACACAT 1980
Db 2029 CCAACTGATCAAGCATCGAGCAGGCTTGAACTGCTTAGCTACTTCATTTACACAT 2088
QY 1981 CCCATCTCATTTCTATCTGCTGAAACATTTGAATCATTTACTGCTGTCTGAAAATGAT 2040
Db 2089 CCCATCTCATTTCTATCTGCTGAAACATTTGAATCATTTACTGCTGTCTGAAAATGAT 2148
QY 2041 GATGAAAAAGTAGCAGAGCTGCACTACAAATTTTCAAAAACACAGAGAACAAATTTGAA 2100
Db 2149 GATGAAAAAGTAGCAGAGCTGCACTACAAATTTTCAAAAACACAGAGAACAAATTTGAA 2208
QY 2101 GAGGATTTTCCACACATCGATCAGCTGGCTTCCGTTTACATACAAAATCTTAAAAA 2160
Db 2209 GAGGATTTTCCACACATCGATCAGCTGGCTTCCGTTTACATACAAAATCTTAAAAA 2268
QY 2161 GGACCCCCCGTCAAGCCAAATATGCGATTCATTTGATCCATGCGATATTTTCTAGTAA 2220
Db 2269 GGACCCCCCGTCAAGCCAAATATGCGATTCATTTGATCCATGCGATATTTTCTAGTAA 2328
QY 2221 GAGACCCAGTTTGACACAGATATTTGAGCTCTGCAATAGAGCTAGATCCAAAGCACTG 2280
Db 2329 GAGACCCAGTTTGACACAGATATTTGAGCTCTGCAATAGAGCTAGATCCAAAGCACTG 2388

QY 2281 GAACATCTCATACACACATTTGTTACTATTGATGATATTTGCTCTCCGACCTGATCAA 2340
Db 2389 GAACATCTCATACACACATTTGTTACTATTGATGATATTTGCTCTCCGACCTGATCAA 2448
QY 2341 TTTGCTGCTCCCTTGGAATCTTGGTAGTACTTTCATTTGTAAGATCTTCATGAAAT 2400
Db 2449 TTTGCTGCTCCCTTGGAATCTTGGTAGTACTTTCATTTGTAAGATCTTCATGAAAT 2508
QY 2401 GATGCGCTTCAGGAAAAAGACACTTAACTTTGGGTTCCAGATGAAGATATCTCCT 2460
Db 2509 GATGCGCTTCAGGAAAAAGACACTTAACTTTGGGTTCCAGATGAAGATATCTCCT 2568
QY 2461 GAGACAAATGGTCAAAATTCAGCTATTAAAAATGATGTCGATGAGCTACTTGAATGAAA 2520
Db 2569 GAGACAAATGGTCAAAATTCAGCTATTAAAAATGATGTCGATGAGCTACTTGAATGAAA 2628
QY 2521 AATAATCACAGTAAATCAGAACTTCTACCTTAAGATTCCTAACACAAATATTCATAGT 2580
Db 2629 AATAATCACAGTAAATCAGAACTTCTACCTTAAGATTCCTAACACAAATATTCATAGT 2688
QY 2581 GATGAGACTTGACAGAACAGGGAAAAATTAGTAAACAGATATGTCAGCTTGACACTT 2640
Db 2689 GATGAGACTTGACAGAACAGGGAAAAATTAGTAAACAGATATGTCAGCTTGACACTT 2748
QY 2641 GCTGCTGGGAGTCTATTGTGAAGCTGGCAACAAGCCGTTACGATCAATCATACACA 2700
Db 2749 GCTGCTGGGAGTCTATTGTGAAGCTGGCAACAAGCCGTTACGATCAATCATACACA 2808
QY 2701 TTAGAACAATATACGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAAGCA 2760
Db 2809 TTAGAACAATATACGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAAGCA 2868
QY 2761 GTGTTTGCCAGAACTTCAAAAGGCTTTCCGTTTACGCTTCCACTTGAGTATATG 2820
Db 2869 GTGTTTGCCAGAACTTCAAAAGGCTTTCCGTTTACGCTTCCACTTGAGTATATG 2928
QY 2821 GCATCTGCGCCCTTTGTGCAAAAGATCCGTTAAAGAGAGAAAGACTCATGCTAGAGCA 2880
Db 2929 GCATCTGCGCCCTTTGTGCAAAAGATCCGTTAAAGAGAGAAAGACTCATGCTAGAGCA 2988
QY 2881 TGTGTTGTAAGAAATTAATATGTAAGCGGAGTATCTGAAGCAGATGCAGCTGTAGT 2940
Db 2989 TGTGTTGTAAGAAATTAATATGTAAGCGGAGTATCTGAAGCAGATGCAGCTGTAGT 3048
QY 2941 GAAAAATTATGTCTCTTACCAAGTATGTTGTTCCATATCAATTCACCTTTGGCA 3000
Db 3049 GAAAAATTATGTCTCTTACCAAGTATGTTGTTCCATATCAATTCACCTTTGGCA 3108
QY 3001 CATGACCCAGATTAATGTCAAAGTACAGATATTTGAACAACCTTAAAGATGTAAAGATGT 3060
Db 3109 CATGACCCAGATTAATGTCAAAGTACAGATATTTGAACAACCTTAAAGATGTAAAGATGT 3168
QY 3061 CTTTGCTTGTCTGAAAAATTAATATGCTTAAAAATGAATTAACAGTACGCTTTTATC 3120
Db 3169 CTTTGCTTGTCTGAAAAATTAATATGCTTAAAAATGAATTAACAGTACGCTTTTATC 3228
QY 3121 AGAAGATGCTAGAAAAATTTTAACAACAAGAAAGATGCCAAGGACCAATGATGCAAAA 3180
Db 3229 AGAAGATGCTAGAAAAATTTTAACAACAAGAAAGATGCCAAGGACCAATGATGCAAAA 3288
QY 3181 ATGAATGAANAACGTGACACTGCTGATGATGTTGCCATGAATATCTCATGTCAAAAGAT 3240
Db 3289 ATGAATGAANAACGTGACACTGCTGATGATGTTGCCATGAATATCTCATGTCAAAAGAT 3348
QY 3241 ACTACATACAGTTTGGAAATCTCTTAAGACCCGGTACTACAGCTGTTCTTCACTCAA 3300
Db 3349 ACTACATACAGTTTGGAAATCTCTTAAGACCCGGTACTACAGCTGTTCTTCACTCAA 3408
QY 3301 CCTGACAAAGATTTTCAAGTAAACCAAAATTAATCTGCTCTGGAATGAAATCATTTTTC 3360
Db 3409 CCTGACAAAGATTTTCAAGTAAACCAAAATTAATCTGCTCTGGAATGAAATCATTTTTC 3468

Qy	3361	ACTCCTGGAAAAACCTTAAACCAACCAATGTTCTAGGAGCTGTTAACACGACCTTTGATCA	3420		
Db	3469	ACTCCTGGAAAAACCTTAAACCAACCAATGTTCTAGGAGCTGTTAACACGACCTTTGATCA	3528		
Qy	3421	GCAGGCAAGCAATCTCAGACCAAAATCATCAGCAATGGAATCTGATGCAATGCAAGCAGC	3480		
Db	3529	GCAGGCAAGCAATCTCAGACCAAAATCATCAGCAATGGAATGGAATCTGATGCAAGCAGC	3588		
Qy	3481	AGCTCAATTCGAAGTCTCTCTGGAAGAAATTAAGGGGAGGCTTGATAGTTCGGAATGAT	3540		
Db	3589	AGCTCAATTCGAAGTCTCTCTGGAAGAAATTAAGGGGAGGCTTGATAGTTCGGAATGAT	3648		
Qy	3541	CACAGTGAATTAAGATTAACAATGTTCTTACCTTCCGGGGAAAAAAGTGACAG	3600		
Db	3649	CACAGTGAATTAAGATTAACAATGTTCTTACCTTCCGGGGAAAAAAGTGACAG	3708		
Qy	3601	AGAGACGACTGATCTTGTAAAGCTGTGAATTTGAGAGCCTAGAGGAGGAAAAAAGC	3660		
Db	3709	AGAGACGACTGATCTTGTAAAGCTGTGAATTTGAGAGCCTAGAGGAGGAAAAAAGC	3768		
Qy	3661	CCCGTCAAGAACAGAGAGGAAATTAAGTATGATGACTTGAATGTTGATACAGAA	3720		
Db	3769	CCCGTCAAGAACAGAGAGGAAATTAAGTATGATGACTTGAATGTTGATACAGAA	3828		
Qy	3721	CAGAAACTTAAGGAGTCAAGCGAAGTCGGAAGAGGCGCATACGCTTCAGAAATCTGAT	3780		
Db	3829	CAGAAACTTAAGGAGTCAAGCGAAGTCGGAAGAGGCGCATACGCTTCAGAAATCTGAT	3888		
Qy	3781	GAACAGCAGTGGCTTGAGGAAAAAGAGCTCAAAAGATATTTAGAAAAATGAAGATGA	3840		
Db	3889	GAACAGCAGTGGCTTGAGGAAAAAGAGCTCAAAAGATATTTAGAAAAATGAAGATGA	3948		
Qy	3841	CAGAAATAGTCCGCCAAAAAAGGTAAAGAGGCGCACCAAAACCTCTTGTTGAGAGT	3900		
Db	3949	CAGAAATAGTCCGCCAAAAAAGGTAAAGAGGCGCACCAAAACCTCTTGTTGAGAGT	4008		
Qy	3901	ACACCAAAAGAGAGCCACAATGAAAATCTTCTAAAAAAGAGCAAAAAAATCTGGA	3960		
Db	4009	ACACCAAAAGAGAGCCACAATGAAAATCTTCTAAAAAAGAGCAAAAAAATCTGGA	4068		
Qy	3961	CCTCCAGACACAGAGGAGGAGGAAGAAAGAAACAAATGGAATTCGGAACAGAG	4020		
Db	4069	CCTCCAGACACAGAGGAGGAGGAAGAAAGAAACAAATGGAATTCGGAACAGAG	4128		
Qy	4021	TCCAAAAGCAACAGCAGCAGTGTCAAGAGACACAGCAGAGAGAAATCTCTGAA	4080		
Db	4129	TCCAAAAGCAACAGCAGCAGTGTCAAGAGACACAGCAGAGAGAAATCTCTGAA	4188		
Qy	4081	TCTAGTGAATTTGAATCCACAGCTCCACACCAAGAAAGAGAGCAATCAAAA	4140		
Db	4189	TCTAGTGAATTTGAATCCACAGCTCCACACCAAGAAAGAGAGCAATCAAAA	4248		
Qy	4141	ACGCCATCACCATCAACAACCAAAAAAATGCT	4172		
Db	4249	ACGCCATCACCATCAACAACCAAAAAAATGCT	4280		
RESULT 2	HS26H231	7444 bp	mrna	linear	PRI 25-JAN-2000
LOCUS	HS26H231				
DEFINITION	Novel human gene mapping to chromosome 13.				
ACCESSION	AL137201				
VERSION	AL137201.1	GI:6759511			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 7444)				
AUTHORS	Rhodes, S. and Huckle, E.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-JAN-2000) E-mail contact: humquerry@sanger.ac.uk				
COMMENT	This cDNA sequence was assembled from public domain ESTs and single				

FEATURES	source
1..7444	location/Qualifiers
1..7444	/organism="Homo sapiens"
/mol_type="mrna"	
/db_xref="taxon:9606"	
/chromosome="13"	
/map="13q12.3"	
1..107	/number=1
108..234	/number=2
127..4470	/number=3
127..4470	/codon_start=1
/product="hypothetical protein"	
/protein_id="CAB69911.1"	
/db_xref="GI:6759512"	
/translation="MAHKTPTNDGKITPPGKVEISDKISKEEMVRRLKMYKTFMD MDQSEEEKELYMLALHLSDFLEKHPDKDVRILVACLIADIRIYAPAPYSPDK LKDIEMFTIRLKGLEDKSPQFNRYFLINIMVMSYINCFELEDSENEIFQLYRK LFSVINNHONKVMHMYDLMSSITCEGDPYSOBLDTQVLVNLVPAKRNLMKQVILA KALKRTQALIEPYITTFNFOYLMAGCTSIDSEHFDLILEYINDSHLLVYLRO LEFKLSNDNERLQVVKLAKMFGAKDSELSAONKPLMOCYLGDFNDIHPILIECY KFASHCLMNHDPDLAKDLLEYLKVRSHPDEELIRIDYIVISYITAAKDLILVNDHLNF VRERTLDRKMRVREKEMMGLQIYKVALQSAAGKDAKQIAWIKDKLHLHYONSID DRLIVERFAOYVPHNLETERMKCYLYATQDLDNAVKRLNMKCONLIRROVLD LLDLIKOPKTDASVKALFESKMYVITRNLPDGRKODEKKRTOYLEDDEKIRKOLEVL VSPSCSCQAQACVREITKTKGNRKQPIPNPLEIKELILEIAYHIDTESISLIRQ VNSTIDGDADEDSVPTDQAIRAGLELKLSTHPIISFSAETFSLLCLAMDE KVAERALQIFKNTGSLIEDPRLRSALPLVLIHRSKSGPRQAKVALHCHIAFSSK ETOFAOIEPTPLHKSIDPSNLEHLITPLVTIGHILALPDQFAAPLSVATFIYKDL MNDRLPGRKTKLWVPDEVEPETNWKIOALIKMVRMLLGMKNHNSKSTLRLTTL ILSHDGDLTEQGIKSPDMSRLRLAAGSAIYKLAQOPEYHEIITLLEOYLOCALA INDE CYQVROYFAQKLIHGLSLRPLPLEYMAICALCAADPKYERAAHAROCILYKINIRREY LKQHAANVSEKLSLLEPYVPTTTHLHAHDVYKVOIDEOLKVRBCLMWLIVLEILMA KNNNSHAIFIRKWEINIKOTKDAOGPDAKKNELLYYCDVAMNITNSKSTYSLESP KDPLYPAEFQIPKFNFSNTKNYLPPEMKSFPTGKPTTYVNLGAVMPLSSASAKQSO TKSSEMETVSNASSNSPSPRIKGRLDSEMDSEDEDYTNSSPLVPGKSKDRDSD DLSVSELEKPRGRKTPYTBQEEKLGMDDLKLVIOPOKPKGSOBSRRGRHARSDDO OMPEBKRLKEDIENEDBQNSPPPKKRGKRGPPKPLGGCTPREPEPTMTSKGSKKSG PPAPEEERERQSGNTEQKSKSKOHRSVRAAQRAEPSSALSTOSTPQKGRGR SKTPSPSQPKKNVGRSKQATKENDSSEEDVFOGSSPVDIIPQERTEEEVSTVA VRRSAKREER"	
235..438	/number=3
439..525	/number=4
526..623	/number=5
624..750	/number=6
751..831	/number=7
832..972	/number=8
973..1088	/number=9
1089..1183	/number=10
1184..1329	/number=11
1330..1481	/number=12
1482..1595	/number=13
1596..1677	/number=14

```

exon          /number-14
              1678..1726
              /number-15
exon          1727..1866
              /number-16
misc_feature  1866..2046
              /note="matches EST AA687846 from clone IMAGE:1219676"
misc_feature  1866..1941
              /note="matches EST AA687776 from clone IMAGE:1219628"
exon          1867..1982
              /number-17
exon          1983..2088
              /number-18
exon          2089..2249
              /number-19
exon          2250..2373
              /number-20
exon          2374..2532
              /number-21
exon          2533..2601
              /number-22
exon          2602..2738
              /number-23
exon          2739..2862
              /number-24
exon          2863..3067
              /number-25
misc_feature  complement(2912..3302)
              /note="matches EST AA630616 from clone 856164"
exon          3068..3182
              /number-26
exon          3183..3315
              /number-27
misc_feature  3302..3572
              /note="matches EST AA333118"
exon          3316..3435
              /number-28
misc_feature  3349..3818
              /note="matches EST AA282190 from clone IMAGE:704631"
exon          3436..3498
              /number-29
exon          3499..3644
              /number-30
misc_feature  complement(3586..4065)
              /note="matches EST A1655429 from clone IMAGE:2239331"
exon          3645..3750
              /number-31
exon          3751..4190
              /number-32
misc_feature  complement(3864..4286)
              /note="matches EST A1680124 from clone IMAGE:2264521"
misc_feature  4001..4119
              /note="matches EST N83765 from clone KK2745"
misc_feature  4135..4450
              /note="matches EST F12354 from clone c-38h11."
exon          4191..4298
              /number-33
misc_feature  4255..4307
              /note="matches EST T64867 from clone 22264"
exon          4299..4434
              /number-34
misc_feature  join(4398..4533,4518..4660,4659..4675,4674..4693)
              /note="matches EST H44781 from clone 188032"
exon          4435..7444
              /number-35
misc_feature  4537..5022
              /note="matches EST A1681422 from clone IMAGE:2272632"
misc_feature  4537..4892
              /note="matches EST AA911125 from clone IMAGE:1416862"
misc_feature  4538..4969
              /note="matches EST A1775866 from clone IMAGE:1877482"
misc_feature  4538..4947
              /note="matches EST AA987361 from clone IMAGE:1603350"

```

```

misc_feature  4538..4943
              /note="matches EST A1299646 from clone IMAGE:1898010"
misc_feature  4562..5003
              /note="matches EST AA447433 from clone 784555"
misc_feature  4571..4854
              /note="matches EST H93424 from clone 220700"
misc_feature  join(4573..4649,4650..4815)
              /note="matches EST W20070 from clone 306049"
misc_feature  complement(4629..4937)
              /note="matches EST AW089775 from clone IMAGE:2593146"
misc_feature  4746..4994
              /note="matches EST AA747568 from clone IMAGE:1269696"
misc_feature  complement(4803..5332)
              /note="matches EST A1911784 from clone IMAGE:2329873"
misc_feature  complement(4828..5296)
              /note="matches EST AA563884 from clone IMAGE:1013990"
misc_feature  4861..5243
              /note="matches EST D20082 from clone pml283."
misc_feature  join(4872..4921,5281..5371)
              /note="matches EST AA177347 from clone 621810"
misc_feature  complement(join(4872..4921,5281..5332))
              /note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature  4872..4921
              /note="matches EST AA154932 from clone 599194"
misc_feature  complement(4872..4921)
              /note="matches EST AW048929 from clone
UT-M-BH1-amo-C-09-0-UT"
misc_feature  complement(4876..5333)
              /note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature  complement(join(4922..4975,5281..5332))
              /note="matches EST A1012649 from clone RPLBA02"
misc_feature  complement(4922..4975)
              /note="matches EST AA899586 from clone
UT-R-E0-cy-e-02-0-UT"
misc_feature  matches EST A1172172 from clone RMBV82"
              complement(4931..5332)
              /note="matches EST AA65961 from clone 745478"
misc_feature  complement(4951..5332)
              /note="matches EST A1140790 from clone IMAGE:1566222"
misc_feature  complement(5021..5309)

```

```

Query Match          99.8%; Score 4165.6; DB 9; Length 7444;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 418; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  ATGGCTCATTCAAAGCTAGGACCAATGATGCAAAATTCATATCCGCTGGGCTCAAG 60
        |||||||
DB      127 ATGGCTCATTCAAAGCTAGGACCAATGATGCAAAATTCATATCCGCTGGGCTCAAG 186

QY      61  GAAATATCGATTAATATCTAAGAGGATGTCGACGATTAAGATGCTTGGA 120
        |||||||
DB      187 GAAATATCGATTAATATCTAAGAGGATGTCGACGATTAAGATGCTTGGA 246

QY      121 ACTTTATGATATGAGCAGCAGCTCTGAAGAAAGAGAGCTTTATTTAAACTAGCT 180
        |||||||
DB      247 ACTTTATGATATGAGCAGCAGCTCTGAAGAAAGAGAGCTTTATTTAAACTAGCT 306

QY      181 TTACATCTTGCTTCACATTTTTCACAGCATCTCGTGAAGATGTTGGCTTACGCTA 240
        |||||||
DB      307 TTACATCTTGCTTCACATTTTTCACAGCATCTCGTGAAGATGTTGGCTTACGCTA 366

QY      241 GCCTGCTGCTTCGATATTTTTCAGAGTTTATGCTCGAAGCTCCTTACACATCCCT 300
        |||||||
DB      367 GCCTGCTGCTTCGATATTTTTCAGAGTTTATGCTCGAAGCTCCTTACACATCCCT 426

QY      301 GATTAACCTAAGAGATATATTATGTTATTAACAAGACAGTGAAGGGCTAGAGATACA 360
        |||||||
DB      427 GATTAACCTAAGAGATATATTATGTTATTAACAAGACAGTGAAGGGCTAGAGATACA 486

QY      361 AAGAGCCCAATTCATATAGGATATTTTATTTACTTGGAACAATTCCTTGCGTCAAGTCA 420
        |||||||
DB      487 AAGAGCCCAATTCATATAGGATATTTTATTTACTTGGAACAATTCCTTGCGTCAAGTCA 546

```


QY	421	TATAACAATATGCTTTGAGTTGAAGATATGCAATGAAATTTTCCACCAGCTATACAGAAC	480
Db	547	TATAACAATATGCTTTGAGTTGAAGATATGCAATGAAATTTTCCACCAGCTATACAGAAC	606
QY	481	TTATTTTACGTTATTAACAATATGGCCACAAATAGAAAGTCCATATGCAATATGCTTACAGCTT	540
Db	607	TTATTTTACGTTATTAACAATATGGCCACAAATAGAAAGTCCATATGCAATATGCTTACAGCTT	666
QY	541	ATGAGCTCTATTTATTTTGAAGAGTATACAGTGTCTCAGAGCTTTTGGATACGGTTTAA	600
Db	667	ATGAGCTCTATTTATTTTGAAGAGTATACAGTGTCTCAGAGCTTTTGGATACGGTTTAA	726
QY	601	GTAATATCTGGTACTGCTCATTAAGAATTTAAACAAGCAGCATATGATTTGGCAAGGCT	660
Db	727	GTAATATCTGGTACTGCTCATTAAGAATTTAAACAAGCAGCATATGATTTGGCAAGGCT	786
QY	661	TTACGAGAGAGACAGCTCAAGCTATTTAGCCATATATATACAGTTTATTAACAGGTT	720
Db	787	TTACTGAGAGAGACAGCTCAAGCTATTTAGCCATATATATACAGTTTATTAACAGGTT	846
QY	721	CTGATGCTTTGGGAAAAACATCTATCAGCGATTTGTACAGACATGTCCTTATTAATTTTG	780
Db	847	CTGATGCTTTGGGAAAAACATCTATCAGCGATTTGTACAGACATGTCCTTATTAATTTTG	906
QY	781	GAGCTCTACATATTTGATAGTACATTTGCTGCTCTCTGTTTAAACCCAGCTTGAATTTAA	840
Db	907	GAGCTCTACATATTTGATAGTACATTTGCTGCTCTCTGTTTAAACCCAGCTTGAATTTAA	966
QY	841	TTAAAGACCAATGATATGAGAGAGCGGCTACAGTGTGTTAAACATACAGCAAAATGTTT	900
Db	967	TTAAAGACCAATGATATGAGAGAGCGGCTACAGTGTGTTAAACATACAGCAAAATGTTT	1026
QY	901	GGGGCAAAAGATTCAGAAATTTGGCTTCTCAAAACAAGCCATTTGGCAGTGTACTTTGGC	960
Db	1027	GGGGCAAAAGATTCAGAAATTTGGCTTCTCAAAACAAGCCATTTGGCAGTGTACTTTGGC	1086
QY	961	AGCTTATATGATATTCATGTCACCAATCCGCTGGAAATGTGGAATTTGCTAGCCATTTGT	1020
Db	1087	AGCTTATATGATATTCATGTCACCAATCCGCTGGAAATGTGGAATTTGCTAGCCATTTGT	1146
QY	1021	CTCATGAACCAATCCGATTTTAGCAAAACCTTAACAGATGTCCTTAAAGTAGAGCCATAT	1080
Db	1147	CTCATGAACCAATCCGATTTTAGCAAAACCTTAACAGATGTCCTTAAAGTAGAGCCATAT	1206
QY	1081	GACCCGAGAGAGCTATTAGACATGATGTTATTTGTCTCAAAAGTTACAGCTCTAAAAG	1140
Db	1207	GACCCGAGAGAGCTATTAGACATGATGTTATTTGTCTCAAAAGTTACAGCTCTAAAAG	1266
QY	1141	GATATTTCTTTGGTCAATGATCCTTACTTAATTTTGTGAGAGAGACAATTTAGACAA	1200
Db	1267	GATATTTCTTTGGTCAATGATCCTTACTTAATTTTGTGAGAGAGACAATTTAGACAA	1326
QY	1201	CGATGGAAGTAGCGAAAGGAAGCCATGATGGGACTGGCCAAATTTATTAACAAATATGCT	1260
Db	1327	CGATGGAAGTAGCGAAAGGAAGCCATGATGGGACTGGCCAAATTTATTAACAAATATGCT	1386
QY	1261	TTACAGTCAGAGCTGGAAGAAAGATGCTGCCAAACAAGATAGCATGGATCAAGACAAATTG	1320
Db	1387	TTACAGTCAGAGCTGGAAGAAAGATGCTGCCAAACAAGATAGCATGGATCAAGACAAATTG	1446
QY	1321	CTACATATATTTATCAAAATAGTATGATGATGACTACTTGTGTAACGAGATCTTGGCT	1380
Db	1447	CTACATATATTTATTTATCAAAATAGTATGATGATGACTACTTGTGTAACGAGATCTTGGCT	1506
QY	1381	CAATACATATGTTCTCCAAATTTTGAACCTACAGACGATGAATGCTTATATTAATCTTG	1440
Db	1507	CAATACATATGTTCTCCAAATTTTGAACCTACAGACGATGAATGCTTATATTAATCTTG	1566
QY	1441	TATGCAACATCGAATTTAAATGCTGTGAAGACATTTGAATGAAATGTGGAATTTGCAAAAT	1500
Db	1567	TATGCAACATCGAATTTAAATGCTGTGTGAAGACATTTGAATGAAATGTGGAATTTGCAAAAT	1626
QY	1501	CTGCTCCGACATCAAGTTAAAGATTTGCTTGACTTGATTAGCAACCCCAAAACGATGCC	1560

Db	1627	CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAAGCAACCCAAACGATGCC	1686
QY	1561	AGTGTCGAAGCCATATTTTCAAAGTATGGTTATTACAAGAAATTTACCTGATCCGTG	1620
Db	1667	AGTGTCGAAGCCATATTTTCAAAGTATGGTTATTACAAGAAATTTACCTGATCCGTG	1746
QY	1621	AAGGCTCGAGATTTTCATGAAAGAAATTCACACAGGTGTTAGAAATGATGAGAAATVAGA	1680
Db	1747	AAGGCTCGAGATTTTCATGAAAGAAATTCACACAGGTGTTAGAAATGATGAGAAATVAGA	1806
QY	1661	AAGCAGTTAGAAAGTACTTGTAAGTCCAAACATGCTCCGCAAGCAGCGCTGAAGGTGTGTG	1740
Db	1807	AAGCAGTTAGAAAGTACTTGTAAGTCCAAACATGCTCCGCAAGCAGCGCTGAAGGTGTGTG	1866
QY	1741	CGTGAAGATTAATAGAAGTTGGGCAACCCCAACAGCTACAAATTCCTTTCTGGAATG	1800
Db	1867	CGTGAAGATTAATAGAAGTTGGGCAACCCCAACAGCTACAAATTCCTTTCTGGAATG	1926
QY	1801	ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCATATAGATACCGAATCTATCAGTGT	1860
Db	1927	ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCATATAGATACCGAATCTATCAGTGT	1986
QY	1861	CTTATTAAACAAGTGAACAAATTCATATAGATGGAACACAGATGATGAAATGAGGGTGT	1920
Db	1987	CTTATTAAACAAGTGAACAAATTCATATAGATGGAACACAGATGATGAAATGAGGGTGT	2046
QY	1921	CCAACGTATCAAGCCATCAGAGCAGGTCCTTAACCTGTTAAGGTACCTCATTTACCAT	1980
Db	2047	CCAACGTATCAAGCCATCAGAGCAGGTCCTTAACCTGTTAAGGTACCTCATTTACCAT	2106
QY	1981	CCCATCTCATTTTCATTCTGCTGAACAATTTGAATCATTTACTGCTGCTGTAANAATGGAT	2040
Db	2107	CCCATCTCATTTTCATTCTGCTGTAACAATTTGAATCATTTACTGCTGCTGTAANAATGGAT	2166
QY	2041	GATGAAAAAGTAGACAGAAGCTGCATACAAATTTTCAAAAACACAGGAGCAAAATTGAA	2100
Db	2167	GATGAAAAAGTAGACAGAAGCTGCATACAAATTTTCAAAAACACAGGAGCAAAATTGAA	2226
QY	2101	GAGGATTTTTCACACATCATAGATCAAGCCCTGCTTCTGTTTATCATCAAAATCPAAAAA	2160
Db	2227	GAGGATTTTTCACACATCATAGATCAAGCCCTGCTTCTGTTTATCATCAAAATCPAAAAA	2286
QY	2161	GGACCCCCCGCTCAAGCCAAATATGCCATTCATTGTATTCATGCGATATTTTCTAGTAAA	2220
Db	2287	GGACCCCCCGCTCAAGCCAAATATGCCATTCATTGTATTCATGCGATATTTTCTAGTAAA	2346
QY	2221	GAGACCCAGTTTGGACACAGATATTTTGAGCCTCTGATATAAGGCTTAAATCCAAAGAACCTG	2280
Db	2347	GAGACCCAGTTTGGACACAGATATTTTGAGCCTCTGATATAAGGCTTAAATCCAAAGAACCTG	2406
QY	2281	GAACATGTCATATACACCAATTTGCTACTATGTTGGTGCATATFGCTGCTTGGACCGATCAA	2340
Db	2407	GAACATGTCATATACACCAATTTGCTACTATGTTGGTGCATATFGCTGCTTGGACCGATCAA	2466
QY	2341	TTTGCTGCTCTTGGAATCTTTGGGTAGCTACTTTCATTGTGAAAGATCTTTCATGAAT	2400
Db	2467	TTTGCTGCTCTTGGAATCTTTGGGTAGCTACTTTCATTGTGAAAGATCTTTCATGAAT	2526
QY	2401	GATGGGCTTCCAGGAAAAAGACAACCTAATCTTTGGGTTCCAGATGAAAGATATCTCCT	2460
Db	2527	GATGGGCTTCCAGGAAAAAGACAACCTAATCTTTGGGTTCCAGATGAAAGATATCTCCT	2586
QY	2461	GAGCAATATGTCAAAAATTCAGGCAATTTAAATGATGTTGGTGGATGGCTACCTGGATGAAA	2520
Db	2587	GAGCAATATGTCAAAAATTCAGGCAATTTAAATGATGTTGGTGGATGGCTACCTGGATGAAA	2646
QY	2521	AATATATCACAATTAATCAGAACTTCTACCTTAAGATTTGTAACACAAATATTCAGATG	2580
Db	2647	AATATATCACAATTAATCAGAACTTCTACCTTAAGATTTGTAACACAAATATTCAGATG	2706
QY	2581	GATGAGAGCTTGACAGAAACAGGGAAAAATTAGTAACACAGATATGTCAGCTTGAGACTT	2640

Db 2707 GATGAGACTTGACAGAACGAGGGAATAATAGTAAACACAGATATGTACGCTGAGACTT 2766
OY 2641 GGTGCGGAGGAGTGTATGTGAAGCTGGACAAAGCCCTGTACATGAATATCATCACA 2700
Db 2767 GCTGCTGGGAGTGTATGTGAAGCTGGACAAAGCCCTGTACATGAATATCATCACA 2826
OY 2701 TTAGAACAATATCAGCTATGTGCTATAGCTATCAACGATGAATGCATCAAGTAAGACA 2760
Db 2827 TTAGAACAATATCAGCTATGTGCTATAGCTATCAACGATGAATGCATCAAGTAAGACA 2886
OY 2761 GTGTTTCCCAAGAACTTCACAAAGCCCTTCCCTTTACGGCTTCACCTGAGTATATG 2820
Db 2887 GTGTTTCCCAAGAACTTCACAAAGCCCTTCCCTTTACGGCTTCACCTGAGTATATG 2946
OY 2821 GGAATCTGTGCTTGTGCAAAAAGTCTGTAAAGSAGAAAGCCCTCATGCTAGGCA 2880
Db 2947 GCAATCTGTGCTTGTGCAAAAAGTCTGTAAAGSAGAAAGCCCTCATGCTAGGCA 3006
OY 2881 TGTTTGTGAAAAATATAAATGTAAAGCGGAGTATCTGAGCAGCATGCTGTACT 2940
Db 3007 TGTTTGTGAAAAATATAAATGTAAAGCGGAGTATCTGAGCAGCATGCTGTACT 3066
OY 2941 GAAAAATTTATGTCTCTTACCAAGATATGTTTCCATATCAATTAACCTTTTGGCA 3000
Db 3067 GAAAAATTTATGTCTCTTACCAAGATATGTTTCCATATCAATTAACCTTTTGGCA 3126
OY 3001 CATGACCCAGATTTATGTCAAGATGATTAAGCAACTTAAGATTTAAAGATGT 3060
Db 3127 CATGACCCAGATTTATGTCAAGATGATTAAGCAACTTAAGATTTAAAGATGT 3186
OY 3061 CTTTGGTTTGTCTGAAAAATTTAAATGTGCTAAAAATGAATTAACAGTACGCTTTTATC 3120
Db 3187 CTTTGGTTTGTCTGAAAAATTTAAATGTGCTAAAAATGAATTAACAGTACGCTTTTATC 3246
OY 3121 AGAAAGATGTGTGAATAATTTAAACAACAAAAAGATCCCAAGGACAGATGATGCAAA 3180
Db 3247 AGAAAGATGTGTGAATAATTTAAACAACAAAAAGATCCCAAGGACAGATGATGCAAA 3306
OY 3181 ATGAATGAAAAACTGTACACGTGTGTGATGTTGCCATTAATATCATGTCACAAAGAT 3240
Db 3307 ATGAATGAAAAACTGTACACGTGTGTGATGTTGCCATTAATATCATGTCACAAAGAT 3366
OY 3241 ACTACATACAGTTTGAATCTCTAAGACCCGCTACTACCAAGCTGTTCTTCACTCAA 3300
Db 3367 ACTACATACAGTTTGAATCTCTAAGACCCGCTACTACCAAGCTGTTCTTCACTCAA 3426
OY 3301 CCTGACAAGATTTGATGATACACCAAAAAATTAATCTGCTCGTAATGAATCATTTTTC 3360
Db 3427 CCTGACAAGATTTGATGATACACCAAAAAATTAATCTGCTCGTAATGAATCATTTTTC 3486
OY 3361 ACTCTGAAAAACCTTAACCAACCAATGTTCTAGAGCTGTTTAAACAGCAGCTTTTCACTCA 3420
Db 3487 ACTCTGAAAAACCTTAACCAACCAATGTTCTAGAGCTGTTTAAACAGCAGCTTTTCACTCA 3546
OY 3421 GGAGGCAAGCAATCTCAGACCAAAATCATCAGCAATGGAAGTGAAGCAATGCAAGCAGC 3480
Db 3547 GGAGGCAAGCAATCTCAGACCAAAATCATCAGCAATGGAAGTGAAGCAATGCAAGCAGC 3606
OY 3481 AGCTCAAAATCAAGCTCTCTCGAAGAAATTAAGGGGAGGCTGTATGATTTCTGAATGAT 3540
Db 3607 AGCTCAAAATCAAGCTCTCTCGAAGAAATTAAGGGGAGGCTGTATGATTTCTGAATGAT 3666
OY 3541 CACATGTGAATAATGAAGATTAACCAATGTTTCACTTTCCGGGGAAAAAAGTGCAG 3600
Db 3667 CACATGTGAATAATGAAGATTAACCAATGTTTCACTTTCCGGGGAAAAAAGTGCAG 3726
OY 3601 AGAGAGCACTGTGATCTTTAGAGCTGTAATTTGAGAGAACCTTAAGAGCAGAGAAAAAAGC 3660
Db 3727 AGAGAGCACTGTGATCTTTAGAGCTGTAATTTGAGAGAACCTTAAGAGCAGAGAAAAAAGC 3786
OY 3661 CCCGTCACAGAACAGAGAGAAATTAAGTATGATGACTTACTTAAGTTGGTATACAGAA 3720
Db 3787 CCCGTCACAGAACAGAGAGAAATTAAGTATGATGACTTACTTAAGTTGGTATACAGAA 3846

OY 3721 CAGAAACCTTAAGGCGATGCGAGGAGTGGAAAAAGAGCCATACGGCTTCAGAAATCTGAT 3780
Db 3847 CAGAAACCTTAAGGCGATGCGAGGAGTGGAAAAAGAGCCATACGGCTTCAGAAATCTGAT 3906
OY 3781 GAACAGCACTGGCTGTAGGAAAAAGAGCTCAAAAGAGATATTTAGAAAAATGAATGAA 3840
Db 3907 GAACAGCACTGGCTGTAGGAAAAAGAGCTCAAAAGAGATATTTAGAAAAATGAATGAA 3966
OY 3841 CAGAAATGTCCGCAAAAAAGGTTAAAGAGCCGACACCAAAACCTTTGGTGAAGT 3900
Db 3967 CAGAAATGTCCGCAAAAAAGGTTAAAGAGCCGACACCAAAACCTTTGGTGAAGT 4026
OY 3901 ACACCAAAAGAGAGCAACATGAAATCTGTAAAAAGGAAAGCAAAAAAATCTGGA 3960
Db 4027 ACACCAAAAGAGAGCAACATGAAATCTGTAAAAAGGAAAGCAAAAAAATCTGGA 4086
OY 3961 CCTCCAGCACAGAGAGAGAGAAAGAAAGAAAGCAAAAGTGAATACGGAACAGAG 4020
Db 4087 CCTCCAGCACAGAGAGAGAGAAAGAAAGAAAGCAAAAGTGAATACGGAACAGAG 4146
OY 4021 TCCAAAGCAAAACAGACCGAGTGTCAAGAGAGCAGACAGAGACAAATCTCTGAA 4080
Db 4147 TCCAAAGCAAAACAGACCGAGTGTCAAGAGAGCAGACAGAGACAAATCTCTGAA 4206
OY 4081 TCTAGTCAATTTGAATCCACAGATGTCACACGACACAGAAAGAGAGAGACCATCAAA 4140
Db 4207 TCTAGTCAATTTGAATCCACAGATGTCACACGACACAGAAAGAGAGAGACCATCAAA 4266
OY 4141 ACGCATCACATCACACCAACCAAAAAAATGT 4172
Db 4267 ACGCATCACATCACACCAACCAAAAAAATGT 4298
RESULT 3
HS095825 5253 bp mRNA linear PRI 05-APR-1999
LOCUS
DEFINITION
Human androgen-induced prostate proliferative shutoff associated
protein (AS3) mRNA, complete cds.
VERSION
095825.2 GI:4559409
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.
Androgen-induced proliferative shutoff in prostate cancer cells
Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
2 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
Early gene expression during androgen-induced inhibition of
proliferation of prostate cancer cells: a new suppressor candidate
on chromosome 13, in the BRCA2-Rb1 locus
J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
JOURNAL
MEDLINE
99229875
PUBMED
10215036
REFERENCE
3 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
Direct Submission
JOURNAL
Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
4 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
Direct Submission
JOURNAL
Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REMARK
COMMENT
On Apr 5, 1999 this sequence version replaced gi:4539617.
FEATURES
1..5253
Location/Qualifiers
/organism="Homo sapiens"

```
gene
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /chromosome="13"
  /map="13q12-q13"
  /cell_line="LNCap"
  /cell_type="cancer cells"
  /tissue_type="prostate"
  /note="exon 1 maps to PI Artificial Chromosome PAC26H23;
  exons 2-24 map to cosmid 267p19; exons 25-34 map to p1
  Artificial Chromosome PAC 49j10"
  1..5253
  /gene="AS3"
  1..46
  /gene="AS3"
  /number=1
  exon
  47..173
  /gene="AS3"
  /number=2
  66..4241
  /gene="AS3"
  /codon_start=1
  /product="androgen-induced prostate proliferative shutoff
  associated protein"
  /protein_id="AAD2134.2"
  /db_xref="GI:4559410"
  /translation="MAHSKTRTNDGKITTPPGYKEISDKISKEEMVRLKMKVKTAMD
  MDSDSEEEKELYNLALHLSDFLKHDPDVRLLVACCLADIPRIYAPRAPYSPDK
  LKDIEMFTRQLGLEDPKSPQFNRVRYLENTIAVWVSYNICELEDSENEFTLYRT
  LFSYINNGHOKYMHMVDLMSSTICGDTVSOQLDVLVNLVPAHRLMKQADLA
  KALKRTAQAIIEPYITNPFNOVLIGKTSISDSEHSDLLLELYNIDSHLLSVLQ
  LEFLKSNDEERLQVVKLAKMGANDSELASQNKPLMOCYLAGFNDIHVPIRECY
  KFASHCLMNHDLAKDLLEYLKVSHDPEERLRHDIVSIYTAARDLLVNDHLEF
  VGERITLDKRMVRKEAMGLAOIYKAYALOSAAKDAKOIAWIKDKLHHYYQNSID
  DLIVERIPAYQVWPHNLFTTERMKCLYYATDILNAVKAINEMKCONILRRQVND
  LLDLIRKPKTDASKAIFSKVMYTRSLPDPGAKQDPMKKTQVLEDEKTRKLEVL
  VSPGSCQABGVREITTKLGNKOPTNPLEIKFLERIAVHIDTESISALIQ
  VNSIDGTADDEGVPTDQIRAGLILNVLASTHPISFSAETPESLACLMDSE
  KYAEALQIIFKNTGSKIIEDEPHIRSALEPLHHSKRGPPQAKYAIHCIAHFSK
  EAQPAQIIEPLHKSIDPSNLHLITPVTIGHIALAPDQPAFLKSLVATFEIVKDL
  MNDRLPKGTKTKLVNPDSEVSPETMVKIOAIKMMVRWMLGKNNHSGSTRLTLT
  ILHSDGLTEOGKISKPDMSRLRAGSAIYKLAQPCYHETITLBOYOCALAINDE
  CYQRYQYFAQKIKHGLSRLLPLEYMAICACADPYKERRAHRQCLVKNTNREY
  LKQHAAYSEKLSLLEPYVVPYTHLADHDYKVDIEQLKDKCECLWVLEILMA
  KNNNSHAFIRKVENIKQTKDAQDPDAKKNELVYCDVANNIIMSKSTYSELSP
  KDPVLPARFTQDPKNFSNTKNYLPQMKSEFTGKRPKTVLGAVNKPLSSAQKQ
  TKSSRMETVGNASSSNPSPGRIGKRLDSEMDHSEMDYTMSSPLPGKSDKRD
  DIVRSELEKPRGRKTPVTEEEKIGMDLLTKIYOBQRPQSORRKGHTASRSDQ
  QMPBEKRLKEDIENEDQNSPPKKKGKRGPRPKLAGSGCTPKEEPTMKTSKGSKKKS
  PPAPEEEEEROSGNTQKSKQKHVSRRAQRASSSESSALESTQSTPQKGRPR
  SKTPSPQPKNV"
  174..377
  /gene="AS3"
  /number=3
  378..464
  /gene="AS3"
  /number=4
  465..562
  /gene="AS3"
  /number=5
  563..689
  /gene="AS3"
  /number=6
  690..770
  /gene="AS3"
  /number=7
  771..911
  /gene="AS3"
  /number=8
  912..1027
  /gene="AS3"
  /number=9
  1028..1122
  /gene="AS3"
  exon
  /number=10
  1123..1268
  /gene="AS3"
  /number=11
  1269..1420
  /gene="AS3"
  /number=12
  1421..1534
  /gene="AS3"
  /number=13
  1535..1616
  /gene="AS3"
  /number=14
  1617..1665
  /gene="AS3"
  /number=15
  1666..1805
  /gene="AS3"
  /number=16
  1806..1921
  /gene="AS3"
  /number=17
  1922..2027
  /gene="AS3"
  /number=18
  2028..2188
  /gene="AS3"
  /number=19
  2189..2312
  /gene="AS3"
  /number=20
  2313..2471
  /gene="AS3"
  /number=21
  2472..2540
  /gene="AS3"
  /number=22
  2541..2677
  /gene="AS3"
  /number=23
  2678..2801
  /gene="AS3"
  /number=24
  2802..3006
  /gene="AS3"
  /number=25
  3007..3121
  /gene="AS3"
  /number=26
  3122..3254
  /gene="AS3"
  /number=27
  3255..3374
  /gene="AS3"
  /number=28
  3375..3437
  /gene="AS3"
  /number=29
  3438..3583
  /gene="AS3"
  /number=30
  3584..3689
  /gene="AS3"
  /number=31
  3690..4129
  /gene="AS3"
  /number=32
  4130..4354
  /gene="AS3"
  /number=33
  4355..4810
  /gene="AS3"
  /number=34
```

BASE COUNT 1755 a 944 c 1074 g 1480 t
ORIGIN
Query Match 99.28; Score 4141; DB 9; Length 5253;
Best Local Similarity 99.58; Pred. No. 0;
Matches 4153; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
OY 1 ATGGCTCATTTCAAGAGCTAGAGCAATGATGCAAAAAATTACATATCCGCTGGGCTCAAG 60
Db ATGGCTCATTTCAAGAGCTAGAGCAATGATGCAAAAAATTACATATCCGCTGGGCTCAAG 125
OY 61 GAAATATCAGATTAATAATCTAAGAGAGAGATGAGACGATTAAGAATGTTGTGAAA 120
Db GAAATATCAGATTAATAATCTAAGAGAGAGATGAGACGATTAAGAATGTTGTGAAA 185
OY 121 ACTTTTATGATATGACAGCAAGCTCTGAAAGAAAAGAGCTTTATTTAAACCTAGCT 180
Db ACTTTTATGATATGACAGCAAGCTCTGAAAGAAAAGAGCTTTATTTAAACCTAGCT 245
OY 181 TTACATCTTGCCTTCAGATTTTCTTCAGACATCTGTTAAAGATGTTGCTTACTGTGA 240
Db TTACATCTTGCCTTCAGATTTTCTTCAGACATCTGTTAAAGATGTTGCTTACTGTGA 305
OY 241 GCCTGCTGCTTGCCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 300
Db GCCTGCTGCTTGCCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 365
OY 301 GATTAACATAAGATATATTTATTTATTAACAAGACAGTGTGAAGGGCTAGAGATACA 360
Db GATTAACATAAGATATATTTATTTATTTATTAACAAGACAGTGTGAAGGGCTAGAGATACA 425
OY 361 AAGAGCCCAATTCATATGATATTTTATTTACTGAGACATTCCTTGGGCTCAAGTCA 420
Db AAGAGCCCAATTCATATGATATTTTATTTACTGAGACATTCCTTGGGCTCAAGTCA 485
OY 426 AAGAGCCCAATTCATATGATATTTTATTTACTGAGACATTCCTTGGGCTCAAGTCA 485
Db AAGAGCCCAATTCATATGATATTTTATTTACTGAGACATTCCTTGGGCTCAAGTCA 545
OY 421 TATTAACATATGCTTTGAGTTAGAGATGACATGAATTTTCAACCAGCTATACAGAAC 480
Db TATTAACATATGCTTTGAGTTAGAGATGACATGAATTTTCAACCAGCTATACAGAAC 545
OY 481 TTATTTTCAGTTTATTAACAATGAGCCACATCAAGAAAGTCCATATGACATGATGACCTT 540
Db TTATTTTCAGTTTATTAACAATGAGCCACATCAAGAAAGTCCATATGACATGATGACCTT 605
OY 546 TTATTTTCAGTTTATTAACAATGAGCCACATCAAGAAAGTCCATATGACATGATGACCTT 605
Db TTATTTTCAGTTTATTAACAATGAGCCACATCAAGAAAGTCCATATGACATGATGACCTT 660
OY 541 ATGACCTCTATTTATTTGTAAGGTGATACAGTGTCTCAGAGCTTTTGATACGGTTTA 600
Db ATGACCTCTATTTATTTGTAAGGTGATACAGTGTCTCAGAGCTTTTGATACGGTTTA 665
OY 601 GTAATATCTGTAACCTGCTCATAAGATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
Db GTAATATCTGTAACCTGCTCATAAGATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 725
OY 661 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
OY 721 CTGATGCTTGGGAAACAATCTATCAGAGATTTGTGAGAGCATGTCTTGTACTTAATTTTG 780
Db CTGATGCTTGGGAAACAATCTATCAGAGATTTGTGAGAGCATGTCTTGTACTTAATTTTG 845
OY 781 GAGCTCTACAAATATTTGATAGTCAATTTGCTCTCTGTTTACCCCAAGCTTGAATTTTAA 840
Db GAGCTCTACAAATATTTGATAGTCAATTTGCTCTCTGTTTACCCCAAGCTTGAATTTTAA 905
OY 841 TTAAGAGCAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db TTAAGAGCAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
OY 901 GGGGCAAGAGATTCAGATTTGCTCTCAAAAACAAGCACTTGGCAGTCTACTTGGGC 960
Db GGGGCAAGAGATTCAGATTTGCTCTCAAAAACAAGCACTTGGCAGTCTACTTGGGC 1025
OY 961 AGGTTTAATGATATTCATGTAACAATCCGCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
Db AGGTTTAATGATATTCATGTAACAATCCGCTGGAATGTGTGAAATTTGCTAGCCATTGT 1085

Db 1026 AGGTTTAATGATATTCATGTAACAATCCGCTGGAATGTGTGAAATTTGCTAGCCATTGT 1085
OY 1021 CTCATGACCATCCGATTTAGCAAAAAGACTTAACAGAGATATCTTAAAGTGGCTACAT 1080
Db CTCATGACCATCCGATTTAGCAAAAAGACTTAACAGAGATATCTTAAAGTGGCTACAT 1145
OY 1081 GACCCCTGAGAGAGCTTATGAGATGATGTTATTTGTCTCAATAGTTTACAGCTCTTAAAG 1140
Db GACCCCTGAGAGAGCTTATGAGATGATGTTATTTGTCTCAATAGTTTACAGCTCTTAAAG 1205
OY 1141 GATATTTCTTGTCTCAATGATCACTTAAATTTTGTGAGAGAGAGAGAGAGAGAGAG 1200
Db GATATTTCTTGTCTCAATGATCACTTAAATTTTGTGAGAGAGAGAGAGAGAGAGAG 1265
OY 1201 CGATGAGAGCTACGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db CGATGAGAGCTACGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
OY 1261 TTACAGTCAAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db TTACAGTCAAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385
OY 1321 CTACATATATATTTATCAAAAATGATATGATGATGATGATGATGATGATGATGATG 1380
Db CTACATATATATTTATCAAAAATGATATGATGATGATGATGATGATGATGATGATG 1445
OY 1381 CAATCATGCTTCTCCACAAATTTAGAACTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db CAATCATGCTTCTCCACAAATTTAGAACTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
OY 1441 TATGCCACACTGATTTAAATGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db TATGCCACACTGATTTAAATGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565
OY 1501 CTGCTCCGACATCAAGTAAGAGATTTGCTTGAATTTGATTAAGCAACCAACAGATGCC 1560
Db CTGCTCCGACATCAAGTAAGAGATTTGCTTGAATTTGATTAAGCAACCAACAGATGCC 1625
OY 1561 AGTGTCAAGGCAATATTTCAAAAAGTATGATATTTACAAAGAAATTTACTGATCTGCT 1620
Db AGTGTCAAGGCAATATTTCAAAAAGTATGATATTTACAAAGAAATTTACTGATCTGCT 1685
OY 1621 AAGGCTCAGAGATTTCAAGAAATTTCAACAGAGTGTGAGAGATGATGAGAAATTAAGA 1680
Db AAGGCTCAGAGATTTCAAGAAATTTCAACAGAGTGTGAGAGATGATGAGAAATTAAGA 1745
OY 1686 AAGGCTCAGAGATTTCAAGAAATTTCAACAGAGTGTGAGAGATGATGAGAAATTAAGA 1745
Db AAGGCTCAGAGATTTCAAGAAATTTCAACAGAGTGTGAGAGATGATGAGAAATTAAGA 1805
OY 1741 CGTGAATTAACATAAGAGTTGGGCAACCCCAACAGCCCTACAAATCTTCTGGAGAAAG 1800
Db CGTGAATTAACATAAGAGTTGGGCAACCCCAACAGCCCTACAAATCTTCTGGAGAAAG 1865
OY 1801 ATCAAGATTTCTTGGAGAGAGATGACACCTGTGACATATGATATGATATGATATGAT 1860
Db ATCAAGATTTCTTGGAGAGAGATGACACCTGTGACATATGATATGATATGATATGAT 1925
OY 1861 CTTATTTAAACAAGTAACAATCAATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db CTTATTTAAACAAGTAACAATCAATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1985
OY 1921 CCAAGTATCAAGCCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db CCAAGTATCAAGCCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2045
OY 1981 CCCATCTCATTTCTGCTGTAAGACATTTGAATATTTACTGCTTGTGTAATTTGAT 2040
Db CCCATCTCATTTCTGCTGTAAGACATTTGAATATTTACTGCTTGTGTAATTTGAT 2105
OY 2041 GATGAAAAAGTAGAGAGAGCTGCACTACAAATTTTCAAAAACACAGAGAGAGAGAGAG 2100
Db GATGAAAAAGTAGAGAGAGCTGCACTACAAATTTTCAAAAACACAGAGAGAGAGAGAG 2165

OY	2101	GAGGATTTTTCACACATCAGATCAGCCTTGCTGCTGTTTACATCACAATCTAAAAA	2160
Db	2166	GAGGATTTTTCACACATCAGATCAGCCTTGCTGCTGTTTACATCACAATCTAAAAA	2225
OY	2161	GGACCCCCCCTCAGGCCAATATAGCCATTGATGTATCCATWGGAAATTTTCTAGTAAA	2220
Db	2226	GGACCCCCCCTCAGGCCAATATAGCCATTGATGTATCCATWGGAAATTTTCTAGTAAA	2285
OY	2221	GAGACCCAGTTTGCACAGATATTTTGAGCCTTGCAATAAGCCTAGATCCAGAACCTG	2280
Db	2286	GAGCCCAAGTTTGCACAGATATTTTGAGCCTTGCAATAAGCCTAGATCCAGAACCTG	2345
OY	2281	GAAGATCTCATAAACACATTTGGTACTATTTGGTCATATTTGGTCCCTGGACCTGATCAA	2340
Db	2346	GAAGATCTCATAAACACATTTGGTACTATTTGGTCCCTGGACCTGATCAA	2405
OY	2341	TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTTGTGAAAGATCTTCTATGAAT	2400
Db	2406	TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTTGTGAAAGATCTTCTATGAAT	2465
OY	2401	GATGGGCTTCCAGGGAAAAAGACAACATACTTTGGGTCCAGATGGAAGATATCTCCCT	2460
Db	2466	GATGGGCTTCCAGGGAAAAAGACAACATACTTTGGGTCCAGATGGAAGATATCTCCCT	2525
OY	2461	GAGACAATGTCAAAATTTCAGGCTATTTAAAAATGATGGTTCGATGGCTACTTGGATGAAA	2520
Db	2526	GAGACAATGTCAAAATTTCAGGCTATTTAAAAATGATGGTTCGATGGCTACTTGGATGAAA	2585
OY	2521	AATATATCACAATAAATCAGAACTTCTACCTTAAGATTTGCTAACACAATATTCATATGT	2580
Db	2586	AATATATCACAATAAATCAGAACTTCTACCTTAAGATTTGCTAACACAATATTCATATGT	2645
OY	2581	GATGAGACTTGCACAGAACAGGGAAAAATTGTAAACAGATATGTCCAGCTGAGACTT	2640
Db	2646	GATGAGACTTGCACAGAACAGGGAAAAATTGTAAACAGATATGTCCAGCTGAGACTT	2705
OY	2641	GCTGCTGGGAATGCTATTTGTGAAGCTGGCACAGAACCTGTATACCATGAATCATCA	2700
Db	2706	GCTGCTGGGAATGCTATTTGTGAAGCTGGCACAGAACCTGTATACCATGAATCATCA	2765
OY	2701	TTTAAACAATATCAGCTATGTGCAATTTGGCTATCAGACGATGAAATGCTATCAAGTAAACAA	2760
Db	2766	TTTAAACAATATCAGCTATGTGCAATTTGGCTATCAGACGATGAAATGCTATCAAGTAAACAA	2825
OY	2761	GTGTTTGGCCAGAAACTTCACAAAAGGCTTCCCGGTAAAGCGCTTCCACTTGAGTATATG	2820
Db	2826	GTGTTTGGCCAGAAACTTCACAAAAGGCTTCCCGGTAAAGCGCTTCCACTTGAGTATATG	2885
OY	2821	GCATCTGTGCCCCCTTTGTGCAAAAGATCTCTTAAAGGAGAGAAAGCTCATGTAGGCAA	2880
Db	2886	GCATCTGTGCCCCCTTTGTGCAAAAGATCTCTTAAAGGAGAGAAAGCTCATGTAGGCAA	2945
OY	2881	TGTTTGGGAAAAATATTAATGTAAAGCGGAGATGTCGAAGCGAGCATGCGCTGTATG	2940
Db	2946	TGTTTGGGAAAAATATTAATGTAAAGCGGAGATGTCGAAGCGAGCATGCGCTGTATG	3005
OY	2941	GAAAAATATTTGCTCTTTCACACAGAGATGTTGTCATATTAACAATTCACCTTTGGCA	3000
Db	3006	GAAAAATATTTGCTCTTTCACACAGAGATGTTGTCATATTAACAATTCACCTTTGGCA	3065
OY	3001	CATACCCAGATTATGTCCAAAGTACAGAGATTTGACACACTTAAAGATGTTAAAGATGT	3060
Db	3066	CATACCCAGATTATGTCCAAAGTACAGAGATTTGACACACTTAAAGATGTTAAAGATGT	3125
OY	3061	CTTTGGTTTGTCTCGAAAAATATTAATGTGCTAAAAATGAAATTAACAGTACAGCTTTTATC	3120
Db	3126	CTTTGGTTTGTCTCGAAAAATATTAATGTGCTAAAAATGAAATTAACAGTACAGCTTTTATC	3185
OY	3121	AGAAAAGATGTAGAAAAATATTTAAACAAACAAAGATGCCAAGACCATGATGATGCAAAA	3180
Db	3186	AGAAAAGATGTAGAAAAATATTTAAACAAACAAAGATGCCAAGACCATGATGATGCAAAA	3245

QY	3181	TTGAATGAAAAAGCTACACGTGTGGAGATGGTGGCATGATATATCATGCTCAAGAGT	3240
Db	3246	ATGATATGAAAACTGTACACTGTGTGATTTGGCATGATATATCATGCTCAAGAGT	3305
QY	3241	ACTACATACAGTTTGGAAATCTCTAAAGACCCGGTACTACAGGCTCGTTCTTCATCAA	3300
Db	3306	ACTACATACAGTTTGGAAATCTCTAAAGACCCGGTACTACAGGCTCGTTCTTCATCAA	3365
QY	3301	CCTGACAGAAATTTTCAGTAACACCAAAATTAATCTGGCTCCTGGAATGAATTCATTTTTC	3360
Db	3366	CCTGACAGAAATTTTCAGTAACACCAAAATTAATCTGGCTCCTGGAATGAATTCATTTTTC	3425
QY	3361	ACTCCTGGAAAACCTTAAACCAACCAATGTTCTAGAGCTGTTTAAACAAGCCACTTTCATCA	3420
Db	3426	ACTCCTGGAAAACCTTAAACCAACCAATGTTCTAGAGCTGTTTAAACAAGCCACTTTCATCA	3485
QY	3421	GCAGGCAAGCAATCTCAGACCAATCATCAGAAATGGAACCTGTAAAGCATGCAAGACG	3480
Db	3486	GCAGGCAAGCAATCTCAGACCAATCATCAGAAATGGAACCTGTAAAGCATGCAAGACG	3545
QY	3481	AGCTCAATTCGAAGCTCTCTGGAAGATTAAGGGGAGGCGCTGTATAGTTCTGGAATGGAT	3540
Db	3546	AGCTCAATTCGAAGCTCTCTGGAAGATTAAGGGGAGGCGCTGTATAGTTCTGGAATGGAT	3605
QY	3541	CACAGTAAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGAAAAAAATGTGACAA	3600
Db	3606	CACAGTAAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGAAAAAAATGTGACAA	3665
QY	3601	AGAGACGCTCTGATCTTGTAAAGCTGTGAATTTGAGAGAAAGCTTAAAGCAGGAAAAAAAG	3660
Db	3666	AGAGACGCTCTGATCTTGTAAAGCTGTGAATTTGAGAGAAAGCTTAAAGCAGGAAAAAAAG	3725
QY	3661	CCCGTACAGAAACGAGGAGAGAAATTAGGTATGATGATCATTAAGTTGGTACAGGAA	3720
Db	3726	CCCGTACAGAAACGAGGAGAGAAATTAGGTATGATGATCATTAAGTTGGTACAGGAA	3785
QY	3721	CAGAAACCTAAAGGCAGTCAAGCAAGTCTGAAAAAGAGCCATTAAGGCTTCAGATCTGAT	3780
Db	3786	CAGAAACCTAAAGGCAGTCAAGCAAGTCTGAAAAAGAGCCATTAAGGCTTCAGATCTGAT	3845
QY	3781	GAACAGCAGTGGCTGAGAGAAAAGAGGCTCAAGAGAGATATATTAGAAAAATGAAGATGAA	3840
Db	3846	GAACAGCAGTGGCTGAGAGAAAAGAGGCTCAAGAGAGATATATTAGAAAAATGAAGATGAA	3905
QY	3901	ACACCAAAAGAAAGCCCAACATGAAACTTTAAAAAAGGAAGCAAAAAAATATCTGGA	3960
Db	3966	ACACCAAAAGAAAGCCCAACATGAAACTTTAAAAAAGGAAGCAAAAAAATATCTGGA	4025
QY	3961	CCTCCAGCACAGAGAGAGAGAAAGAAAGAACCAAAAGTGGAAATATCGGAACAGAG	4020
Db	4026	CCTCCAGCACAGAGAGAGAGAAAGAAAGAACCAAAAGTGGAAATATCGGAACAGAG	4085
QY	4021	TCCAAAAGCAAAACGACCGCAGGTGTCAAGAGAGACACGACGAGAGCAAGATCTTCGAA	4080
Db	4086	TCCAAAAGCAAAACGACCGCAGGTGTCAAGAGAGACACGACGAGAGCAAGATCTTCGAA	4145
QY	4081	TCTAGTGAATTTGATCCACACAGTCCACACACACAGAAAGAGACGGAAGACATTCAGAA	4140
Db	4146	TCTAGTGAATTTGATCCACACAGTCCACACACACAGAAAGAGACGGAAGACATTCAGAA	4205
QY	4141	ACGCCATCACCATCACAACCAAAAAAATATGTG	4173
Db	4206	ACGCCATCACCATCACAACCAAAAAAATATGTG	4238

associated protein AS3 mRNA, complete cds.
AY102267
VERSION
AY102267.1 GI:31321922
KEYWORDS
SOURCE
ORGANISM
Mus musculus
(house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
1 (bases 1 to 5317)
Geck, P., Maffini, M., Sonnenschein, C. and Soto, A.M.
The AS3 proliferative arrest gene has an ancient eukaryotic
heritage and shows highly conserved functional domains in mice
Proc. Ann. Meet. Am. Assoc. Cancer Res. 43, 987 (2002)
REFERENCE
AUTHORS
TITLE
2 (bases 1 to 5317)
Geck, P., Maffini, M., Liang, S.L., Sonnenschein, C. and Soto, A.M.
Extreme domain conservation and expression of the mouse AS3
proliferative arrest protein
Unpublished
JOURNAL
REFERENCE
AUTHORS
TITLE
3 (bases 1 to 5317)
Geck, P., Maffini, M., Liang, S.L., Sonnenschein, C. and Soto, A.M.
Direct Submission
Submitted (07-MAY-2002) Anatomy and Cell Biology, Tufts University
School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA
FEATURES
source
1..5317
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="5"
/map="B4CM; G2; between BRCA2 and KL"
/sex="male"
/tissue_type="brain"
147..4487
/codon_start=1
/product="androgen-induced prostate proliferative shutoff
associated protein AS3"
/protein_id="AA052216.1"
/db_xref="GI:31321923"
/translation="MAHSKTRINDGKITYPGCVKEISDKISKSEENVRLKMYKTFMD
MDQSEEREKELNLAHLASDFLKHDPKDVRLVACCLADIFRIYAEAYTSPDK
LKDIFMTROLKLEDRSPQFNRYFLLENIAWKSYNICEFELEDSEIFTOLERT
LFSVINGNCHQVHMHVDLMSSIIPEGDTVOELDLVNLVMPHAKRLNNOADLA
KALKRTAOAIEPYITNFENOVLMIGKTSISDSEHVDLITLIEVINDSILLISVLP
LEFKLSDNERLDYVYKLLAMFGAKSELSQNKPLMOCTIGRNDIHYVRIECLY
KRASHCLMHDPALADTEYLVKVRSHDEPALARHDTVSIVPAVKDILLVDHLIN
VRERTLDRWRVRKEMMGLAQIYKRYSLQSAAGDAAOISWAKDKILLVDHLIN
DLILVERIPAOYVPHNLLETTERMKCLYLYATLIDNAKALINEMKCCNLRHQED
LIDLIRKQRTDASVKAIFSRYVITRNLDPEKADPFMKFTQVLEDDERIKROLEAL
VSPSCSKOAEQCVREITRKLGPNROPTPEPLEMTKELIERIAPYHIDESIILIKO
VVKSIDGTADDEGVPTDOAIRAGLELLKLVSPTRHPSFHEAEPESTLACKMDE
KYAEALQIFKRTGSKIEDPPIHSALLPVJHHKSKPPPOAKAIHICHTAITSK
ETQFAQIFEPRLKSLDPSNLEHLITPVYIGHIALAPQFAAPLKSIVATIVLDL
EHLIDRPGKRTKLWVDEVSPEYVTKIOAIKMYRWMLGKMNHNSKSTSLRLIT
ILHSDGLTEQOKISKPDMSRLIAAGSAIVKLAQEPCHYEITTEQOYLALAINDE
CYOVROFAOKLHKISRLRLPLEYMAICALCAPVKERRAHAROCLEKNTVIREY
LKOHASVEKLISLLEPYVYPTIHLADPXYVODIEOLKDKYECQFVLEIEM
KNEHNSHAFIRKMYENIKOTKDAQCPDDTKKMEKLYTVCDVAMNIIKMSSTYSLESP
KDPVLPAREFTQPDKNFSNTKNTLPEPKMSFTTPGKPTANVLGAVNPKLSAGKQSO
TKSSHMETVSNASSSNPSGRIGRLDSSMDHSENDYMASSDLPKSKDKREDP
DLSELEKPRSKKAPVTDPEEKIGMDLTKLVEQOKPKSGRGRKGRASDSQDM
PEERKHKEELLENEDEONSPPKGRGRPKPLGGSTKEETPMKTSKGRKTKLVP
VYDDDEEEROIGMTEHKSQSKQHTSKRAOORASPEPSVNESQSTPKGRGRSK
APSPQPKKIVGRSKOYATREKENDSEEMDVLOASSPVSDTTOBGAEEDISGVN
RRRSKRER"

BASE COUNT 1676 a 1100 c 1199 g 1342 t
ORIGIN

Query Match 83.2% Score 3471.8 DB 10; Length 5317;
Best Local Similarity 89.7% Pred. No. 0;
Matches 3741; Conservative 0; Mismatch 422; Indels 6; Gaps 1;
QY 1 ATGGCTCATTTCAAGACTAGACCAATGAGAAAATTCATATCCGCTGGGTCAAG 60

|||||
147 ATGGCTCATTTCAAGAGCAGGACCAACGATGGGAAAATTCATACCTCCTGGAGTCAAG 206
QY 61 GAAATATTCAGATPAAATATCTAAGAGAGATGCTGAGACGATTAAGATGCTGTGAAA 120
|||||
207 GAAATCTCAGATPAAATCTCTAAGAGAGATGCTGAGCGGTGTAAGATGCTGTGAAA 266
QY 121 ACTTTATGATATGAGACCAAGACCTGTGAAGAGAAAAGAGCTTATTTAAACCTAGCT 180
|||||
267 ACTTTCATGACATGAGCAGCAGACCTGTGAAGAGAAAAGAACTTATCTAACCCTAGCT 326
Db 181 TTACATCTTGCTTCAGATTTTTTTTTCAGACATCTCGTAAGAGATGCTGCTACAGTGA 240
|||||
327 TTACATCTTGCTTCAGATCTCTCTCCCAAGCATCTCGATTAAGATGCTGCTTACAGGAG 386
QY 241 GCGTGTGCTGCTGCTGATATTTTTCAGAGATTTATGCTCTGAAAGCTCTTACACATCCCT 300
|||||
387 GCTTGCTGCTGCTGCTGATATTTTTCAGAGATTTATGCTCTGAGGCTCTTACAGTCTCCC 446
QY 301 GATTAACATAAGATATATTTATGTTTATTAACAAGACAGTTGAAGGGCTAGAGATPACA 360
|||||
447 GATTAACATAAGATATATTTATGTTTATTAACAAGACAGCTAAGGAGCTAGAGATPACA 506
Db 361 AAGAGCCCAAAATTCATAGATATTTTATTTACTTGAGAGACATTTGCTGGGTCAAGTCA 420
|||||
507 AAGAGCCCTCAGTTATATAGTATTTTATTTACTTGAGAGACATTTGCTGGGTCAAGTCA 566
QY 421 TATTAACATATGCTTTGAGTTAGAGATAGCAATGAATTTTCAACCAGCTATPACAGAAC 480
|||||
567 TATTAACATATGCTTTGAGTTAGAGATAGCAATGAATTTTCAACCTTATPACAGAAC 626
Db 481 TTATTTTCAGTTATTAACAATGCGCCCATTCAGAAAGTCCATATGACATGGAAGCTT 540
|||||
627 TTATTTCTCAGTTATTAACAATGCGCCCATTCAGAAAGTCCATATGACATGGAAGCTT 686
QY 541 ATGAGCTATATTTATTTTGTGAGGTGATPACAGTGTCTCAGAGACTTTTGGATACGTTTA 600
|||||
687 ATGAGCTATATTTTGTGAGGTGATPACAGTGTACAGTGTCTCAGAGACTTTTGGATACGTTTA 746
Db 747 GTAATCTGCTGCTACGCTCCCATATCAAGCTTAACCAAGCAAGCATATGATTTGGCAAGGCT 806
QY 601 GTAATCTGCTGCTACGCTCATAGAATTTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
|||||
807 TTGCTGAAGAGAGCTGCTCAACCTATGAGCAATATATTAACCAATTTTAAACAGGTT 866
Db 721 CTGATGCTTTGGGAAAACATCTATACAGCGATTTTGTGAGACATGCTTTGACTTAATTTTG 780
|||||
867 CTGATGCTTTGGGAAAACATCTATACAGTGTATTTGTGAGACATGCTTTGACTTAATTTTG 926
QY 781 GAGCTCTACATATTTGATAGTCAATTTGCTGCTCTCTGTTTACCCAGCTTGAATTTAA 840
|||||
927 GAACCTCTACAAATGACGACCATATGCTGCTTCTGCTCCGACGCTTGAATTTAA 986
QY 841 TTAAGAGAGATATGATATGATAGGAGGCTACAGGCTGTTAAACTACTGTGCAAAAGTGT 900
|||||
987 TTAAGAGAGATATGATATGATAGGAGGCTTCCAGGTTGTTAAACTACTGTGCAAAAGTGT 1046
Db 901 GGGGCAAAAGATTCAGAAATTTGCTCTCAAAACAGCCACTTTGGCAGTCTACTTTGGC 960
|||||
1047 GGGGCAAAAGATTCAGAAATTTGCTCTCAAAACAGCCACTTTGGCAGTCTACTTTGGC 1106
QY 961 AGCTTTATGATATCTATGATACCAATCCGCTGGAATGTGTGAATTTGCTGACCATTTGT 1020
|||||
1107 AGCTTTATGATATCTATGATACCAATCCGCTGGAATGTGTGAATTTGCTGACCATTTGT 1166
Db 1021 CTCATGAACCATCTGATTTAGCAAAAGACTTAACAGAGATATCTTAAGTGAAGTCAAT 1080
|||||
1167 CTCATGAACCATCTGATTTAGCAAAAGACTTAACAGAGATATCTTAAGTGAAGTCAAT 1226
QY 1081 GACCTTGAGAGAGCTATTTAGACATGATGTTATTTGTCTCAATAGTTACACCTCTTAAAG 1140
|||||

Db 1227 GATCCTGAGAGAGCTATTAGACATGATGTTATGTGCTATATGATPAACAGCTGCTAAAAAG 1286
Qy 1141 GATATCTCTTGTGCAATGATGACCTTACTTAATTTTGTGAGAGAGAGAACATTTAGACAAA 1200
Db 1287 GATATCTCTTGTGCAATGATGACCTTACTTAATTTTGTGAGAGAGAGAACATTTAGACAAA 1346
Qy 1201 CGATGAGAGATGAGCAAAAGAACCGATGATGGAGCTTGGCCAAATTTATAGAAATATGCT 1260
Db 1347 CGGTGAGAGATGATGCAAAAGAACCGATGATGGAGCTTGGCCAAATTTATAGAAATATGCT 1406
Qy 1261 TTACAGTACGAGCTGGAAGAAAGATGCTGCAAAACAGATAGATGATGATCAAGACAAATG 1320
Db 1407 TTACAGTACGAGCTGGAAGAAAGATGCTGCAAAACAGATATCTGGGTCAAGGACAACTG 1466
Qy 1321 CTACATATATATTTTCAAAATAGTATGATGATGACCTTGTGTAAGCGATCTTGTCT 1380
Db 1467 CTGATATATATCTTCAAGAACAGATGACGACCGGCTACTGTTGAGCGGATCTTGTCT 1526
Qy 1381 CAATACATGCTTCTCACAATTTAGAAACTACAGACGAGATGAAATGCTTATATTTACTTG 1440
Db 1527 CAGTACATGCTCCTCACAATTTAGAAACGACGAGACGATGAAGTGTGCTGATTTACTTG 1586
Qy 1441 TATGCAACATGATTTAAATGCTGTGAAGACATTTGAATGAAATGCTGAAATGCTAAAT 1500
Db 1587 TACGCTACCTGAGCTTGAATGCTGTGAAGACATTTGAATGAAATGCTGAAATGCTAAAT 1646
Qy 1501 CTGCTCGACATCAAGTAAGAGATTTGCTTACTTGTATTAAGCAACCCAAACAGATGCC 1560
Db 1647 CTACTCGACATCAAGTAAGAGATTTACTTGTATTTAATAACAACCAAAACAGATGCC 1706
Qy 1561 AGTGTCAAGGCAATTTTCAAAAGTATGCTTATTAACAGAAATTTACCTGATCTGCT 1620
Db 1707 AGTGTCAAGGCAATTTTCAAAAGTATGCTTATTAACAGAAATTTCCAGATCTGCT 1766
Qy 1521 AAGCTCAGATTTCAAGAAATTCACACAGGCTGAGAAAGATGATGACAAATATAGA 1680
Db 1767 AAGCTCAGATTTCAAGAAATTCACACAGGCTGAGAAAGATGATGACAAATATAGA 1826
Qy 1681 AAGCAGTTAGAGTCTTGTAGTCAACATGCTCTGTCAGACAGGCTGAAGGTTGTTG 1740
Db 1827 AAACAGGTGAAGGCTTGTAGTCCACATGTTCTGTCAGACAGGCTGAAGGCTGTTG 1886
Qy 1741 CGTGAATTAAGTAAGAGTTGGCAACCCCAAGACGCTACAAATCTTCTCTGGAATG 1800
Db 1887 CGTGAATTAAGTAAGAGTTGGCAACCCCAAGACGCTACAAATCTTCTCTGGAATG 1946
Qy 1801 ATCAAGTTCTCTTGGAGAGATGACACCTGTCACATAGATAGATGATGATGATGATGAT 1860
Db 1947 ATCAAGTTCTCTTGGAGAGATGACACCTGTCACATAGATGATGATGATGATGATGAT 2006
Qy 1861 CTTATTTAAACAGTGAACAAATCAATAGATGAACAGCAGATGATGAAGTGAAGGTT 1920
Db 2007 CTTATTTAAACAGTGAACAAATCAATAGATGAACAGCAGATGATGAAGTGAAGGTT 2066
Qy 1921 CCAACTGATCAAGCATCAGCAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1980
Db 2067 CCAACTGATCAAGCATCAGCAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2126
Qy 1981 CCAACTGATCAAGCATCAGCAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2040
Db 2127 CCAACTGATCAAGCATCAGCAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2186
Qy 2041 GATGAAAAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2100
Db 2187 GATGAAAAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2246
Qy 2101 GAGATTTTCCACATCAGATCAGCTGCTGCTGTTTATCATCACAATCTTAAAAA 2160
Db 2247 GAGATTTTCCCTCACAATCAGATCAGCTGCTGCTGTTTATCATCACAATCTTAAAAA 2306
Qy 2161 GGAGCCCCCGCTCAAGCAAAATATGCAATGATGATGATGATGATGATGATGATGATGATGAT 2220
Db 2307 GGAGCCCCCGCTCAAGCAAAATATGCAATGATGATGATGATGATGATGATGATGATGATGAT 2366

Qy 2221 GAGACCCAGTTTGGACAGATATTTGAGCCTCTGATAGAGACCTAGATGCAAGCAACCTG 2280
Db 2367 GAAACCCAGTTTGGACAGATATTTGAGCCTCTGATAGAGACCTAGATGCAAGCAACCTG 2426
Qy 2281 GACATCTCATTAACACATTTGCTTACTATTGCTATATGCTCTCTTGGACCTGATCAA 2340
Db 2427 GAGATCTTATTAACCCCTCTGCTCATTAATGATGATGATGATGATGATGATGATGATGATGAT 2486
Qy 2341 TTTGCTGCTCTTGGAAATCTTGGGTGCTCTTCAATTTGAAAGATCTTCTCATGAT 2400
Db 2487 TTTGCTGCTCTTGGAAATCTTGGGTGCTCTTCAATTTGAAAGATCTTCTCATGAT 2546
Qy 2401 GATGCGCTTCCAGGAAAAAAGACAACTTAACTTTGGTTCAGATGAAAGATTTCTCT 2460
Db 2547 GACCGCTTCCAGGAAAAAAGACAACTTAACTTTGGTTCAGATGAAAGATTTCTCT 2606
Qy 2461 GAGACATGCTCAAAATTTCAAGCTTATTAATGATGCTGATGCTGATGCTGATGATAA 2520
Db 2607 GAGACAAATGCTCAAAATTTCAAGCTTATTAATGATGCTGATGCTGATGATAA 2666
Qy 2521 AATATCAGATTAATATGAGAACTTCTACCTTAAGATGCTTAAACAACAATTTGATAGT 2580
Db 2667 AATATCAGATTAATATGAGAACTTCTACCTTAAGATGCTTAAACAACAATTTGATAGT 2726
Qy 2581 GATGAGACTTGAAGAGAGAGGAAAAATTTAGTAAACAGATATGCTACGCTGAGACTT 2640
Db 2727 GATGAGACTTGAAGAGAGAGGAAAAATTTAGTAAACAGATATGCTACGCTGAGACTT 2786
Qy 2641 GCTGCTGGAATGCTATTTGTGAAGCTGACACAGACCTTTTACATGAATATATCA 2700
Db 2787 GCTGCTGGAATGCTATTTGTGAAGCTGACACAGACCTTTTACATGAATATATCA 2846
Qy 2701 TTAGCAATATATCAAGCTTGTGATTAAGTCAACGATGATGATGATGATGATGATGATGAT 2760
Db 2847 CTGAGCAGTACCAAGCTTGTGATTAAGCTTACATGATGATGATGATGATGATGATGATGAT 2906
Qy 2761 GTGTTGGCCAGAAACCTTACAAAGGCTTTCCGTTTAAAGCTTCCACTGAGTATAT 2820
Db 2907 GTGTTGGCCAGAAACCTTACAAAGGCTTTCCGTTTAAAGCTTCCACTGAGTATAT 2966
Qy 2821 GCAATCTGTGCTTTTGTGCAAAAGATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Db 2967 GCAATCTGTGCTTTTGTGCAAAAGATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3026
Qy 2881 TGTTTGTAAGAAATATTAATGTAAGGCGGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Db 3027 TGTTTGTAAGAAATATTAATGTAAGGCGGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAG 3086
Qy 2941 GAAAAATATTTGCTCTTCTACAGAGATGTTGCTCATTAATTCACCTTTTGGCA 3000
Db 3087 GAAAAATATTTGCTCTTCTACAGAGATGTTGCTCATTAATTCACCTTTTGGCA 3146
Qy 3001 CATGACCCAGATTTATGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060
Db 3147 CATGACCCAGATTTATGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3206
Qy 3061 CTTTGGTTTGTGTAAGATTTAATGCTTAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 3120
Db 3207 CTTTGGTTTGTGTAAGATTTAATGCTTAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 3266
Qy 3121 AGAAAGATGTAAGAAATTTAAACAAACAAAGATGCGCAAGACAGATGATGCAAA 3180
Db 3267 AGAAAGATGTAAGAAATTTAAACAAACAAAGATGCGCAAGACAGATGATGCAAA 3326
Qy 3181 ATGAATGAAAAACGTACACTGTGTGATGTTGCGCATGAATATCATCATCTCAAAAGAGT 3240
Db 3327 ATGAATGAAAAATTTGTAAGAGCTGTGTGATGTTGCGCATGAATATCATCATCTCAAAAGAG 3386
Qy 3241 ACTACATTCAGATTTGGAATCTCTTAAAGACCGGAGTACACAGCGCTTTTCACTCAA 3300
Db 3387 ACCACGTACAGCTTGAAGTCTCTTAAGAGACCCGCTGCTGACAGCTGGTTTTCACCCAG 3446


```

OY 3301 CCTGACAGAAATTCAGTAACACACAAATATATCTGCTCTGAAATGAATCATCTTTC 3360
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3447 CCTGACAGAAATTTTATGTAACACACAAATTTTACCTCTCCGAAATGAATCATCTTTC 3506
OY 3361 ACTCTGAGAAACCTTAAACCAACCAATGTTCTAGAGAGCTGTTAACACCACTTTCATCA 3420
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3507 ACTCTGAGAAACCTTAAACCAACCAATGTTCTAGAGAGCTGTTAACACCACTTTCATCA 3566
OY 3421 GGAGCAAGCAATCTCAGACCAATATACCAATGGAACCTGTAAGCAATGCAACGAC 3480
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3567 GGAGCAAGCAATCTCAGACCAATATCAATGGAACCTGTAAGCAATGCAACGAC 3626
OY 3481 AGCTCAAAATCCAGAGCTCTCTGGAAGAAATGAAGGGAGGCTGATAGTCTGTAATGAT 3540
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3627 AGCTCAAAACCAAGCTCTCTGGAAGCAATGGAAGGGAGGCTGATAGTCTGTAATGAT 3686
OY 3541 CACAGTGAAGAAATGAAGATTACCAATGTTCTTACCTTCCGGGGAAGAAAGTGACAG 3600
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3687 CACAGTGAAGAAATGAAGATTATCAATGTTCTTACCTTCCAGGAAAGAAAGTGACAG 3746
OY 3601 AGAGCACTGATCTGTAAGTCTGTAATGGAAGCAAGCTTGAAGGCAAGGAAAGG 3660
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3747 AGAGCAAGCACTGATCTT-----TCTGAGTTGAGAAAGCTTGAAGGCAAGGAAAGG 3800
OY 3661 CCCGTCACAGAACAGAGAGAGAAATTAGTATGATGACTGACTAAGTTGATACAGAA 3720
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3801 CCTGTCACAGACCCCTGGAAGAGAAATTAGTATGATGACTGACTAAGTTGATACAGAA 3860
OY 3721 CAGAAACCTTAAAGCAGTACGCGAAAGTGCGAAAGAGGCCATACGCGCTTACAGATCTGAT 3780
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3861 CAGAAACCTTAAAGCAGTACGCGAGGCGCGAAAGAGGCCCTTACAGCTTACAGCTCAGAC 3920
OY 3781 GAACAGCACTGAGGCTGAGGAGAAAGGCTCAAGAAATATATGAAATGAAGTGA 3840
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3921 GACGACAGTGGCTGAGGAGAAAGGCGCAAGAGAGAGCTCTCGAATATGAGGTTGAG 3980
OY 3841 CAGAAATAGTCCGCAAAAGAGGTTAAAGAGGCGCGACCAAAACCTTGTGTGAGAGT 3900
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3981 CAGAAAGCCGACCAAAAGAGGCAAGAGGCGCGACCAAAACCTTGTGTGAGAGG 4040
OY 3901 ACACCAAAAGAGAGCCCAATGAAATCTTAAAAAGAAAGCAAAAGAAATCTGGA 3960
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4041 ACATGAGAGAGAGAGCCCAATGAAATCAAGAAAGAGAAAGCAAGAAAAAATCTGTA 4100
OY 3961 CCTCAGCAGCAGAGAGAGAGAGAGAAAGCAAGTGAATACGAAACAGAG 4020
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4101 CCTCTGTATGATGAGAGAGATGAGAGAGAAAGCAATGGAACACAGAACCTAAG 4160
OY 4021 TCCAAAGCAAGCAGCAGAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGATCTCTGAA 4080
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4161 TCAAAAGCAAGCAGCAGAGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGATCTCTGAA 4220
OY 4081 TCTAGTGCATTTGAATTCACACAGTCCACACAGAGAAAGAGAGAGAGAGAGAGAG 4140
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4221 ACAAGTGCAGTTGAATTCACACAGTCCACACAGAGAAAGAGAGAGAGAGAGAGAG 4280
OY 4141 ACGGCATCACTCAACCAACCAAAAGAA 4169
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4281 GCACCATCACTCAACCAACCAACCAAAAG 4309
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

RESULT 5
AK122414 5944 bp mRNA linear ROD 15-MAR-2003
DEFINITION Mus musculus mRNA for mKIAA0979 protein.
ACCESSION AK122414
VERSION AK122414.1 GI:28972559
KEYWORDS FLI CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1

```

```

AUTHORS Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S.,
Nakajima, D., Nagase, T., Ohara, O. and Koga, H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
REFERENCE 2 (bases 1 to 5944)
AUTHORS Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics: 2-6-7
kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: mousekazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
source
1..5944
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbx00002"
/lisue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
1..5944
/gene="mKIAA0979"
<1..3577
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=2
/evidence="not_experimental"
/product="mKIAA0979 protein"
/protein_id="BAC65696.1"
/db_xref="GI:28972560"
/translation="FDLILEYINSHLLSVLPQLEFKNSNDNEPRLOYKILAKM
FGAKSELASQKPLMOCTYLGFNDRHPIRECYKSPASHCLMNPDLAKDTELKLV
RSHDEEATRHQVIVSYVAKKDLILVNDLNLNRETKDLKRRVRKREAAAGLAI
YKYSIQSAGKDAKQISWAKDLILHYONSIDRLIVERIFQVYVPHNMLETER
MKCLYLYATDLNVAKALNEMWKCQNLRHQVNDLILDKPKDAVSKALFYSVM
ITRNLPDGRKADDFPKFTFOVLEDEKIRKQLEALVSPSCSKQAKGVRSTKLG
PROPNPELEMLKELLERAPVHIDTESISALIKOVNSIDGTADDEPGVTDQAIR
AGLEILKLSFTPIPSHAFESBLACKKDDKVAFAAQTIRKNSKTEEPH
IRSALEPVLRHRSKGPPEKQATVILHCHALFSSKETQFAQIEPLRLSLDSNLEHL
ITPLVTIGHILAPDOFAAPLKSVAFTIVDLMLNRLPKRTTKLWVDEVSPE
TWKIOALIKMVRWMLGKNNHSGSTSTRLITLILSDGDLTEQGIKSPDMSRL
LAAGSAIYKLAQEPYHEITLEOYOLALAINDCQYOVROFAOKLHRKLSRLPLT
EYMAICALCARKDPKERRAHAROCILVNTVREYLKQAAVSEKLLSLPEYVVPYT
IHLAHPDYKVVODIEOLKDYKECLMPEYLEILMAKNENNSHAFIRKVENIKQIKDA
OGPDPTKMKNEKLYIVCDVAMNITMSKSTYSLESKPDVLPFRFTQDPKNSNTKNY
LPPEKSFETPKPKPTANVLGAVNKLPSAGKOSQTKSRMCTVNASNSNPSSGR
IKGRDSSSENEEDYMSPLPGKSKDEKEDDLSELRPRSRKAPVDPPEKL
GMDLITLIVQEBQPKGSORGRKRTASDSDOOPPEEKRRHEILLNDEQNSPK
GKRGKRPPLGGSTSKKEPPTMKSKKKKKTKTTPYVVDDEDEEIOGNTWTKSKSKO
HRTSKRAQGRAPSPETSFAVESTQSTPQGRGPRSPAPSPQPKIRGRSKQVATKE
NDSSEEMVQLASSVSDTTQEGAEEDIDVGNVRRRSKRERR"
BASE COUNT 1892 a 1151 c 1251 g 1650 t
ORIGIN
Query Match 67.5%; Score 2816.6; DB 10; Length 5944;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 3045; Conservative 0; Mismatches 354; Indels 6; Gaps 1;
OY 765 CTTTGACTTAATTTGAGAGCTTACAAATTAATGATGCTGCTGCTGCTTTACC 824
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 CTTTGATTTGATCTTGAGAGCTTACAACTTACACACCACTTTCGCTTCTGCTGCGC 60
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
OY 825 CCAGCTGAATTTAAATTAAGAGCAATGATATGAGAGAGCGGCTTACAAAGTTGTTAACT 884
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 CCAGCTTGAGTTTAATTAAGAGCAATGATATGAGAGAGCGGCTTACAAAGTTGTTAACT 120
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

OY	885	ACTGGCAAAATGTTTGGGGCAAAAGATTCAAGATTTGGCTTCTCAAAACAAGCCACTTGT	944
Db	121	ACTGGCGAAATGTTTGGGGCAAAAGATTCAAGATTTGGCTTCTCAAAACAAGCCCTCTGT	180
OY	945	GCAGGCTACTTGGGCGAGTTTAATGATATCCATGTACCAATTCGGCCCGGAATGTGAA	1004
Db	181	GCAGTGCTACTGGGCGAGTTTAATGATATCCATGTACCAATTCGGCCCGGAATGTGTAA	240
OY	1005	ATTTCGATGCCATTGTCTCATGAACCAATCCATGATTTTACAAAGACTTAAACAGATATCT	1064
Db	241	ATTTCGCCACGACCTGGCTCATGAACCAATCCGATTTTACAAAGATTTAAACAGATATCT	300
OY	1065	TAAAGTAGGTGCATATGACCTGAGGAAGCTATTAGACATGATGTTATTGTCTCAATAGT	1124
Db	301	TAAAGTAGGTGCATATGATCCTGAGGAAGCTATTAGACATGATGTTATTGTCTCAATAGT	360
OY	1125	TACAGCTCTTAAAAAGGATATCTTTCGGTCAATGATACCTACTATTTTGTGGAGGA	1184
Db	361	AACAGCTCTTAAAAAGGATATCTTCTTGTCAATGATACCTTACTCAATTTTGTGGAGGA	420
OY	1185	GAGAACATTTAGACAAACGATGAGAGATACGCAAAAGAACCCATGATGGGACTTGGCCAAAT	1244
Db	421	GCGAACATTTAGACAAACGATGAGAGATACGTAAGAAAGCCATGATGGGACTGGCCACAGAT	480
OY	1245	TTATTAAGAAATATCCTTTACATGACGACGCTGGAAAAAGATCTCTCAAAACAGATATGATG	1304
Db	481	TTATTAAGAAATATCCTTTTACATGACGACGCTGGAAAAAGATCTCTCAAAACAGATATCCTG	540
OY	1305	GATCAAAACACAAATTTGCTACATATATATATATCAAAATATGTAATTTGATGCATACTTGT	1364
Db	541	GGTCAAGSACAAAGCTGCTGCATATATACATCAAGACGATTCGACGACCGGCTACTTGT	600
OY	1365	TGAACGATCTTTGCTCATATACATATGTTCTCTACAAATTTAGAAACTACAGAACGGATGAA	1424
Db	601	TGAGGGGATCTTTGCTCATATACATATGTTCTCTCAAAATTTGAAACGACGAAACGGATGAA	660
OY	1425	ATGCTTATATATCTTTGTATGGCAACGAGATTTAAATCCTGTGAAGACATGATGAAT	1484
Db	661	GTCTCTTATATCTTTGTATGGCTATCGTACACGAGCTTGAATCCTGTGAAGACATGATGAAT	720
OY	1485	GTGGAATATGTCMAAATCTGCTCGCACATCAAGTAAAGATTTGCTTACTTGATTAAGCA	1544
Db	721	GTGGAATATGTCMAAATCTACTCCGACATCAAGTAAAGATTTACTGTATTAATTAACA	780
OY	1545	ACCCAAAACATGCCAGTGTCAAGGCCATATTTTCAAAAGTATGTTATTACAGAAA	1604
Db	781	ACCCAAAACATGCCAGTGTCAAGGCCATATTTTCAAAAGTATGTTATTACACAGAAA	840
OY	1605	TTTACCTATCTGCTGTAAGGCTCGAGATTTATGAAAGAAATTCACAGGCTGTAGAAA	1664
Db	841	TTTGCAGATCTCTGTAAGGCTCGAGATTTATGAAAGAAATTCACAGGCTGTAGAAA	900
OY	1665	TGATGAGAAATTAAGAAAAGCAGTTAGAGTACTTGTATGTCCAACATGCTCTCTCAAGCA	1724
Db	901	TGATGAGAAATTAAGAAAAGCAGTTAGAGTACTTGTATGTCCAACATGCTCTCTCAAGCA	960
OY	1725	GGCTGAAGTGTGTGCGTGAATAATCTAAGAAAGTTGGGCAACCCCTAAGCGACCTACAAA	1784
Db	961	GGCTGAAGTGTGTGCGTGAATAATCTAAGAAAGTTGGGCAACCCCTAAGCGACCTACAAA	1020
OY	1785	TCCCTTCCTGGAATGATCAGTCTTCTTGGAGAGGTATGACCTGTCGACATAGATATAC	1844
Db	1021	CCCCTTCTTGGAAATGATCAGATTTCTCTTGGAGAGGTATGCTCTGTACACATAGATATAC	1080
OY	1845	CGAATCTATATCAGTCTCTTATTAACAAGTAMCAAAATCAATAGATGGAACAGACAGATGA	1904
Db	1081	TGAATCCATCAGTCTCTTATCAAGCAAGTAAATTAATCAATAGATGGAACAGACAGATGA	1140
OY	1905	TGAAGTATAGGCTGTCCAACTGATCAAGCTATACAGGCGAGCCCTTGAGCTGTGAAGGT	1964
Db	1141	TGAGATATAGGCTGTCCAACTGATCAAGCTATACAGGCGAGCCCTTGAGCTGTGAAGGT	1200

QY	1965	ACTC	CAATTTACACATCCCACTCAATTCATTCGCTGGAACATTGGAATCAATACGGC	2024
Db	1201	GCTGCTTTTACGCA	CCCTATCTCCCTTACATCTGCTGAGACGTTTATGTCCTCTGGC	1260
QY	2025	TTGTC	GAAGAAATGCATGATGAAAAAGTAGCAGAGCTGCACCTACAAAAATTTTCAAAACAC	2084
Db	1261	TTGTC	TGAAGATGATGATGATGGAAGGTGGCAAGGTGGCAAGCTGCACCTACAAATTTTCAAAACAC	1320
QY	2085	AGGA	AGCAAAATGGAAGGATTTTCCACACATCAGATCAGCCTTGCTCCGTTTACA	2144
Db	1321	AGGA	AGCAAAATGGAAGGATTTTCCCAATCAGATCAGCCTTGCTCCGTTTACA	1380
QY	2145	TCACA	AATCTAAAAAGGACCCCGGCAAGCCAAATATGCCATTAATGATATCAATGC	2204
Db	1381	TCACA	AATCTAAAAAGGACCCCGGCAAGCCAAATATGCCATTAATGATATCAATGC	1440
QY	2205	GATAT	TTTCTATGAAGAGACCAGTTTGCACACATATTTGAGCCTCGCATAGACCCT	2264
Db	1441	CATAT	TTTCTATGAAGAGACCAGTTTGCACACATATTTGAGCCTCGCATAGACCCT	1500
QY	2265	AGAT	CCAGCAACCGGAACATCTCATACCACTGGTTACTTGTGATATGCTCT	2324
Db	1501	AGAT	CCAGCAACCGGAACATCTCATACCACTGGTTACTTGTGATATGCTCT	1560
QY	2325	CCTTG	CACTATCATTTGCTGCTCCCTTGGAAATCTTGGGTACTTCAATTTGAA	2384
Db	1561	TCGTG	CACCTGATCAGTTCCGCTGCTCCCTGAAATCTTGGGTACTTCAATTTGAA	1620
QY	2385	AGAT	CTTCTCATGAAATGATGCGCTTCCAGGAAAAAGCAACTAATCTTGGTTCAGA	2444
Db	1621	GGAC	CTCTCATGAAATGATGCGCTTCCAGGAAAAAGCAACTAATCTTGGTTCAGA	1680
QY	2445	TGA	AAATATCTCCTGAGACAATGTCACAAATTCAGGCTTTAAATGATGGTTCATG	2504
Db	1681	TGAG	AAATCTCCTGAGACAATGTCACAAATTCAGGCTTTAAATGATGGTTCATG	1740
QY	2505	GCTA	CTTGGAATGAAAAATTAATCAACAGTAATCAGAACTTACCTTAAGATTGCTAAC	2564
Db	1741	GCTA	CTTGGAATGAAAAATTAATCAACAGTAATCAGAACTTACCTTAAGATTGCTAAC	1800
QY	2565	AACA	TATGATGATGATGATGAGACCTTGACAGAACAGGGGAAATTTGTAACACAGTAT	2624
Db	1801	AACA	TATGATGATGATGATGAGACCTTGACAGAACAGGGGAAATTTGTAACACAGTAT	1860
QY	2625	GTC	AGCTTGAGACTTGCTGCTGGAGTGCATTTGTGAAGTGGCACAGAACCTTGTTA	2684
Db	1861	GTC	AGCTTGAGACTTGCTGCTGGAGTGCATTTGTGAAGTGGCACAGAACCTTGTTA	1920
QY	2685	CCAT	GAATCATCATATTAGAACAAATATCAGCTATGTGCATTTAGCTTCAACAGTATG	2744
Db	1921	CCAG	AGATCATCATATTAGAACAAATATCAGCTATGTGCATTTAGCTTCAACAGTATG	1980
QY	2745	CTAT	CAATAGACAAGATGTTGGCCAGAAACCTCACAAAGGCGCTTCCCGTTTACGCT	2804
Db	1981	CTAT	CAATAGACAAGATGTTGGCCAGAAACCTCACAAAGGCGCTTCCCGTTTACGCT	2040
QY	2805	TC	CACTGAGATATGGAATCTGTGCTTGTGCAAAAAGATCTGTAAAGAGAGAG	2864
Db	2041	TC	CACTGAGATATGGAATCTGTGCTTGTGCAAAAAGATCTGTAAAGAGAGAG	2100
QY	2865	AG	CTCATGCTGGAAGATGTTGGGAAAAATATTAATGTAAGGGGGAGATCTGGAAGCA	2924
Db	2101	AG	CTCATGCTGGAAGATGTTGGGAAAAATATTAATGTAAGGGGGAGATCTGGAAGCA	2160
QY	2925	GC	ATGACAGCTGTAGTGAATAATTTGTCTTCTACACAGATATTTTCCATATAC	2984
Db	2161	GC	ATGACAGCTGTAGTGAATAATTTGTCTTCTACACAGATATTTTCCATATAC	2220
QY	2985	AAT	TACACTTTTGGCACAATGACCCAGATATGCTAAAGTACAGATATTTGAACAACCTTA	3044
Db	2221	AAT	TACACTTTTGGCACAATGACCCAGATATGCTAAAGTACAGATATTTGAACAACCTTA	2280
QY	3045	AG	ATGTTAAAGAAATGCTTGTGGTTGTGTTGGAATATTTATGGCTTAAATGCAAAATTA	3104

Db 2281 AGATGTGAAGATGGCTTGTGTTGTCGAGATATGATGGCTAAATATGAAAAACA 2340
QY 3105 CAGTCACGCTTTATATGAGAAGATGTTAGAAAATATTAAACAACAAAGATGCCAAG 3164
Db 2341 CAGCCATGATTTATGAGAAAAATGTTAGAAAAATATTAAACAACAAAGATGCTCAAG 2400
QY 3165 ACCAGATGATGCAAAAAATGAAAAACGTACTGTGTGATGTTGCCATGAAATAT 3224
Db 2401 ACCAGATGATGCAAAAAATGAAAAATTTGACACCGTGTGTGATGTTGCCATGAAATAT 2460
QY 3225 CATCATGTCAAAAGACTACTACATAGTTTGAATCTCTCAAAAGACCGGTACTATACAG 3284
Db 2461 CATCATGTCAAAAGACACACATGACGCTGAGTGTCTCTCAAAAGACCGGTACTATACAG 2520
QY 3285 TCGTTCCTTCACTCAACCTGACAGAAATTTCACTAACACCAAAATTTCTGCTCTCGA 3344
Db 2521 TCGGTTTTCACCCACCTGACAAAGAAATTTTAAACACCAAAATTTCTGCTCTCGA 2580
QY 3345 AATGAATCATTTTCTACCTCCGGAACCACTAAACCAACAACTGTCTAGACCTGTAA 3404
Db 2581 AATGAATCATTTTCTACCTCCGGAACCACTAAACCAACAACTGTCTGAGCAGTAA 2640
QY 3405 CAAGCCACTTTCATCAGACGAGCAAGCAATCTCAGACCAATCATCAGAAATGGAATCTG 3464
Db 2641 TAAGCCACTTTCATCAGACGAGCAAGCAAGTCTCAGACCAATCATCAGAAATGGAATCTG 2700
QY 3465 AAGCAATGCACAGCAGCAGCTCAATTCAGCTCTCTCGAAGAAATAAGGGAGCTTGA 3524
Db 2701 GAGCAACGACAGCAGCAGCTCCAAACCAAGCTCTCTCGAAGAAATAAGGGAGCTTGA 2760
QY 3525 TGGTTCGAAATGATCAGAGTAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 3584
Db 2761 TGGTTCGAAATGATCAGAGTAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 2820
QY 3585 GAAAAAAGTGAACAGAGACGACTGTGTTGAGGCTGAAATGGAAGAGCTAG 3644
Db 2821 AAAAAAAGTGAACAGAGAGAACGCTGATCTT-----TCGAGTGTGAGAGAGCTAG 2874
QY 3645 AGGACGAAAAAAGCCCGCTCAGACAGACAGAGAGAAATTAAGTATGATGATGATGATGATG 3704
Db 2875 AAGTCGAAAAAAGCCCTGACAGACCTGAGAGAAATTAAGTATGATGATGATGATGATGATG 2934
QY 3705 TAAGTGTGACAGAGACAGAACTTAAGGACCTGACGAGTGGAAAAAGAGCCATAC 3764
Db 2935 TAAGTGTGACAGAGAGAACTTAAGGACCTGACGAGTGGAAAAAGAGCCGCTAG 2994
QY 3765 GGGTTCAGATGTGATGAACAGCAGTGGCTGAGAGAAAGAGGCTCAAGAGATATATAT 3824
Db 2995 AGCTTCAGACTCAGACAGAGAGAGAGTGGCTGAGAGAAAGAGGACCAAGAGAGAGCTCT 3854
QY 3825 AGAAAAATGAGATGAACAGATATGTCGCCAAAAAAGGGTAAAGAAGCCGACACCAAA 3884
Db 3055 GGAATTTAGAGATGACAGAGAACAGCCACCAAAAAAGGGGAAAAAGAGCGACGACCAAA 3114
QY 3885 ACCCTTGTGAGAGTGAACAGAAACCAAAAGAGCCAAATGATGAAATCTTCTAAAGAGAG 3944
Db 3115 ACCCTTGTGAGAGGAGATGCAAGAGAAAGAGCCCAATGATGAAATCTTCTAAAGAGAGAA 3174
QY 3945 CAAAAAATGATGAGACCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4004
Db 3175 CAAAGAAAAAAGTTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3234
QY 4005 AATATCGAAGACAGAGTCCAAAAAGCAAGCAGCGAGTGTCAAGAGAGACACAGAGAG 4064
Db 3235 AATATCGAAGACATATATGCAAAAAAGCAAGCAGCGACATCAAAAGAGAGACACAGAGAG 3294
QY 4065 AGAGAAATCTGATGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4124
Db 3295 AGAGAAATCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3354
QY 4125 AGGAAAGCAGATCAAAAAAGCAGATCAGATCAGAAACCAAAAAA 4169
|||||

Db 3355 AGGAGAGCAGATCAAAAGCAGCAGATCAGATCAGAAACCCCAAAAA 3399
RESULT 6
AF294791
LOCUS
DEFINITION
Homo sapiens SCC-112 (SCC-112) mRNA, complete cds.
ACCESSION
AF294791
VERSION
AF294791.1
KEYWORDS
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 6744)
Kumar,D., Patel,S., Whiteside,T.L. and Kasid,U.
Identification and characterization of SCC-112, a novel cell cycle
regulated gene in human cancer cells
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 6744)
Kumar,D. and Kasid,U.
Direct Submission
Submitted (09-AUG-2000) Radiation, Medicine, 3970 Reservoir Rd, NW,
Washington, DC 20007, USA
FEATURES
source
1..6744
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="4"
1..6744
/gene="SCC-112"
232..4125
/gene="SCC-112"
/codon_start=1
/product="SCC-112"
/protein_id="AA082347.1"
/db_xref="GI:21951802"
/translation="MIRKRLKVVYKTTMDMDQSEDEKQOYLPLALHASEFRLRNPK
DRLVLAACCLADIFRIYAPAEPTSHDKLKDIFLTTRDLKGLIEDKSPQNFYFL
ENLAWKSYNICEFELEDNEIFIQLEFRTLESYNNHNKRYVMHMDLSSILMEGD
VFOELGSLILNIPAKHKNKOSFDLAVLLKRYOTIEACIANFNQVLYVGRSSV
SPDSEHFDPLIOELFADPHLLSVAPDLFKKSDGERLAVVLLAKLKGSRAD
LATONRPLMOCFLGRNDHIVPRLESYVFASCLMNHBDLAKDLEYIKVSHDPE
ATRHDIYVITTRAKRDALVNDQLEFVBERLDRKVRKAMKGLAQLRKTCILH
GAGKFAAEKSWIKDLHIYQNSIDKLVLEKTFAYLVPHNLETERMKCLYL
VASLDPNVAKALNEMKCOMLRSYHRELLDHPKPTSEANCAMFGKLMATKANLPD
PCKAOPFVKFNQVLDDEKLSOLELLSPSCSKQADICVLEIRKLANPQPTNP
PLEMVKFLERLAPVHIDSEALISALVLLANKSTEGTADDEEGVSDPTAIRGSELLK
VLSFTPTPSFASERYESILQCLRMEDKVAABAIDIPNTGKITEPDIPOIRSTLIP
ILHOKAKRGTPIHOKAVHCHIAIFNKEVLAQIFELRSILNADVPOLITPLVSL
GHISMLAPDQFASPMKSVANFTVNDLNDNRSTGRKNGKLSMPDEVSPEVLAQVA
IKLVLMILGKMNOSKANSSTLRLSAMLVSEGDLEQRISKSDMSRYRLAAGSAI
MKLAOEPCHYEITTPPOCALVIDECYQVREGDLEKHLALVLLPLEYMAIFA
LCAKDPVREBRARHAROCCLKNISIREYIKQPMATEKLSLPLEVYVYVYMLHLAD
POPTRSQVDQDLDRICECLMPELVLMETNENNSAPFKMAENIKLITDAOSPDESK
TNEKLYTCVDVALCVINRSALCNADSPDVPVPMKFTQPERDCNDSYISEBTRV
LLTGKRPKAGVIGAVNKLISATGRKPYRSTGTETGSIINVSLENPSTGRNRSRS
SEAAETGVSENEENPVRILISVTPVKNIIDVKNKEINSDATOGNISDSGKRTVYAA
GNSIDQKTEDEKVEDSGPAPSKPRGRPKSPSOGNATKNDLNPINKGRRAAVG
OESPGSLGKMKAPKLPDLAKKAPAEKQIDLPK"
BASE COUNT 2156 a 1243 c 1400 g 1945 t
ORIGIN
Query Match 43.5%; Score 1815.6; DB 9; Length 6744;
Best local Similarity 69.7%; Pred. No. 0;
Matches 2457; Conservative 0; Mismatches 1069; Indels 0; Gaps 0;
QY 32 GAAAAATTTCATATCCGCTGGGCTCAAGGAATATACATTAATATATTAAGAGGAGA 91
Db 173 GGAAGATCGCTTACCTCCGGGGGTAAAAAGAGATCCCAAGAAAGACACAGAGAGAGA 232
QY 92 TGGTAGAGGATTAAGGATGTTGTGAAAACTTTATGATATGAGCAGAGACTGAGG 151
|||||

Db 233 TGATCAAAAGCGCTGAGATGGTAGGAAACCTTATGATATGATGAGACGACTGAGAG 292
Qy 152 AAGAAAGAGAGCTTATTTAAACCGTATACACTGCTGACAGATTTTTCGACAGC 211
Db 293 ATGAAAAACACAGATATCTCCACAGCTTGACCTTGACATCTGAATTTCTCCACAGA 352
Qy 212 ATCTGTGTAAGATGTTGCTTACTGTAAGCCGTGCTGCTGCTGATATTTTCAGAGATT 271
Db 353 ACCCAATAAAGATGTGGCTCTCTGTGATGATGTTTGGCTGATATCTTTCGTACT 412
Qy 272 ATGCTCTGAAGCTCTTACACATCCCTGATTAACCTAAGAGATATATTTATGTTTAA 331
Db 413 ATGCCCCAGAGCTCCATATCTCTCCATGAATTAAGACATATTTTGTATTA 472
Qy 332 CAGACAGTTGAGGGGCTAGAGGATACAAAGAGCCAAATTCATAGATATTTT 391
Db 473 CCAGACATTAAGAGTTGGAGAGATACAAAGAGCTCCACAGTTTATGATATCTTTAT 532
Qy 392 TACTGAGACATGCTGGGTCAAGTCAATTAACATATGCTTGAAGTTAGAAGATGCA 451
Db 533 TATTAGAAATTTAGCTGGGTAAATCATATTAACATCTGCTTGAATTTGAAGATTGCA 592
Qy 452 ATGAAATTTTCCACCAGTATACAGAACCTTATTTTCAGTTATTAACATGGCCCATC 511
Db 593 ATGAAATTTTATCAGCTTTTGAACCTCTCTCTCAGTATCAACATATGCCCCAATA 652
Qy 512 AGAAAGTCCATATGACATGTTAGACCTTATGAGCTTATTTGTAAGGTGATACAG 571
Db 653 AGAAGTCAAAATGACATGCTAGATTGATGAGTTCATCATGATGAGAGTGTATGAG 712
Qy 572 TGTCTCAGAGCTTTGGATACGGTTTAAATCTGTAATCTGTAACCTGCTCATAGAAATTA 631
Db 713 TTACCAAGAAATTTATGGGCTCCATCTTATTAACCTCATCTCCACATTAAGAACT 772
Qy 632 ACAAGCAGCAATATGTTGGCAAGGCTTACTAATAAGAGACAGCTCAAGCTATGAGC 691
Db 773 ATAAACATCTCTGACCTTGCAAAAAGTCAATTAAGAAAGAACATCCACAGCTATGAGG 832
Qy 692 CATATATTAACCACTTTTATATCAGTTCGATGCTTGGGAAACATCTATCAGCAGATT 751
Db 833 CATGATTTGCCAATTTTTCATCAACAGTCTGCTGCTGGAAGATCATCAGTAGATT 892
Qy 752 TGTCAAGACATGCTTTTACTTAATTTTGGAGCTCTACAAATTTGATGTCATTTGCTGC 811
Db 893 TGTCAAGACATGATTTATCTGATTCAGGAACCTTTTGTCAATATATCTCATATTTAT 952
Qy 812 TCTCTGTTTACCAGCTTGAATTAATTAAGAGCAATGATGAAGAGGCGCTTAC 871
Db 953 TATCGTATGCGACAGCTTGAATTTCAACTTAAGAGCAATGATGAGAGAGCGCATTTAG 1012
Qy 872 AAGTTGTTAACTACTGGCAAAAATGTTGGGGCAAGAGATTCAGAAATTTGGCTTCTCAA 931
Db 1013 CTGTTGTTGACCTTACTTAATGTTTGGGCTCCAAAGATTTCTGATTTGGACAACAGA 1072
Qy 932 ACAAGCCACTTTGGAGTGTACTTGGCAGSTTAAATGATATCCATGACCAATCCGCC 991
Db 1073 ATCGCTCTTTGGGAATGTTTCTTGAGCAGTTTAAATGATATCATGTTCTGTGAAT 1132
Qy 992 TGAATGTGTAATTTGCTAGCCATGTCATCATGATACCATTCGATTTAGCAAAAGCT 1051
Db 1133 TAGAAGTGTGAATTTGCCAGTCACTGTTTAAATACCCACATTTAGGGAAGATC 1192
Qy 1052 TAACAGATATCTTAAAGTGAAGTCAATGACCTTGAGGAAGCTTTAGACATGATGTTA 1111
Db 1193 TCACAGATATTTAAAGTTAGATCAATGATCCAGAAAGAGCTTTTGTCTATGATGICA 1252
Qy 1112 TTGTCTCAATAGTACAGCTGCTAAAAGATATTTCTTGTGTAATGATCATCTTACTTA 1171
Db 1253 TTGTACTATATATACAGCTGCAAGAGAGAGCTGGCTTGTATTAATGATGAGCTGTG 1312
Qy 1172 ATTTGTGAGAGAGAACTTAGCAAAACGATGAGAGTACGCAAGAGCCATGATGG 1231
Db 1312 ATTTTGTGAGAGAGAACTTAGCAAAACGATGAGAGTACGCAAGAGCCATGATGG 1231

Db 1313 GCTTTGTAAAGGAAAGAACTGATATAAACGCTGGCGAGTAAAGAAAGAACCTATGATG 1372
Qy 1232 GACTTGGCCCAATTTATTAAGAAATATGCTTTACAGTCCAGCTGGAAGAAAGATCTGCA 1291
Db 1373 GTCTGGCTCAGCTTTATTAAGAAATATCTCTTCAATGGTGAAGCAGAAAGAGACTGAG 1432
Qy 1292 AACAGATGATGATCAAGCAAAATGCTACATATATATTAATTAATTAATAGTATGATG 1351
Db 1433 AGAAAGTCAAGCTGATAAGAGCAAACTTGTGCAATTTATTAATCAAGAAAGACTTGACG 1492
Qy 1352 ATCGACTCTTGTGAAGGATCTTTGCTCAATATGATGTTCTTCACAAATTTAACAAC 1411
Db 1493 ACAACCTGTTGTAGAGAAATCTTGTGCTAGTATCTTGTCCACACAACTGGAAGAG 1552
Qy 1412 CAGAAAGGATGAATGCTTATATTAATGATGACACACTGATTAATTAATGCTGTAAG 1471
Db 1553 AAGAGAGATTAATGCTTATATTAATTAATGATGATGATTTGATGATGATGATGATG 1612
Qy 1472 CATGATGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1531
Db 1613 CTCTCAAGCAAGATGGAAGTGTAGAAACATGCTCGAGGACATGATGATGATGATG 1672
Qy 1532 ACTGATTAAGCAACCCCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1591
Db 1673 ATTTGACAGACAGCTATACATGATGATGATGATGATGATGATGATGATGATGATG 1732
Qy 1592 TTATTAAGAAATTTTACCTGATCTGTAAGGCTCAGAGATTTGATGAAGAAATTCAC 1651
Db 1733 CCATGCAAGAAATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1792
Qy 1652 AGGTGTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1711
Db 1793 AGGTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1852
Qy 1712 GCTCTGCAAGCAGGCTGAAGTGTGCTGATGATGATGATGATGATGATGATGATGATG 1771
Db 1853 GTTCTGCAAAACAGCAGATTTGTGTGAAGAAATGATGATGATGATGATGATGATGATG 1912
Qy 1772 AACAGCTTCAAAATCTTCTGGAATGATCAATGATGATGATGATGATGATGATGATG 1831
Db 1913 AGCAACCAAAATCTTCTGAGATGATGATGATGATGATGATGATGATGATGATGATG 1972
Qy 1832 TGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1891
Db 1973 TGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2032
Qy 1892 GAACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1951
Db 2033 GGACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2092
Qy 1952 AACTGCTTAAGGTACTCATTTACATCCATCTCATTTTCTGCTGAACATTTG 2011
Db 2093 AACTCTTAAGGTCTGCTTACATCTCTTACATCTCTCTCTCTCTCTCTCTCTCTCT 2152
Qy 2012 AATCATTTACTGCTTGTGTAATGATGATGATGATGATGATGATGATGATGATGATG 2071
Db 2153 AGTCTTGTATCAGTCTGTAAGAAAGGATGATGATGATGATGATGATGATGATGATG 2212
Qy 2072 TTTTCAAAACAGAGAAATTTGAAGAGATTTTCCACATCTGATGATGATGATGATG 2131
Db 2213 TTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2272
Qy 2132 TTCTGTTTACATCAAAATCTTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 2191
Db 2273 TTCCATTTTTCATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 2332
Qy 2192 ATGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2251
Db 2333 ACTGTATACAGGCTTATTAACAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 2392
Qy 2252 TGCAATAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2311
Db 2393 TCAGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2452

OY	2312	GTCAATATGCTCTCCTTGCACCTGATCATTTTGGCTCTTGGAATCTTGGTAGCTA	2311
Db	2453	GCACATTTTCATGCTTTCACACCAATCGTTGGCTTCCCAATGAAATCTGTAGACCA	2512
OY	2372	CTTTCATTTGAAAGATCTCTCATGATGATGCGCTTCCAGGGAAAAAGCACTAAC	2431
Db	2513	ATTTTATTTGTGTAAGATCTGCTAAATGATGACAGGTCAACAGGTGAAGATGGAANAC	2572
OY	2432	TTTGGGTTCCAGATGAAGAACTATCTCCTGAGACAAATGCTAAATTCAGGCTATTAAA	2493
Db	2573	TGTGTCCTCCAGATGAAGAGGTTTCCCTGAAGTACTAGCAAGGTACAGGCATTTAAC	2632
OY	2492	TGATGCTTCCATGGCTACTTGGAAATGAAAATTAATCACAGTAAATCAGAACTTCTACT	2551
Db	2633	TTTGTGTAAGGTGGCTGTGGGTATGAAAAAACAACCACTCTAAATCTCCATTCACACC	2692
OY	2552	TAAATTTGCTAACAAACATATTTCATAGTAGTGAGAGCTTACAGAACAGGGAAAAATTA	2611
Db	2693	TTTGGTTTATTTCAGCCGATGTGGTTAGTAGGGGGGACCTACAGAGAAAAAGAGATCA	2752
OY	2612	GTTAAACCAAGATATGTCACGCTCTGAGACTTGTCTGCGGAGTCTATTGTGAAGCTGCAC	2671
Db	2753	GTAATATCGATATGCTCTGCTTGGGATAGCTGCTGGTAGTGCCATATGAAGCTTGCTC	2812
OY	2672	AAGAACCCCTGTACCATGAATTCATCATTAAGAAACATATCAGTATGCTCATTTAGCTA	2731
Db	2813	AGGAACCTGTTTACCATGAAATTTATTTCACCAGAACACTTTCAGCTGTCTGCACACTGTTA	2872
OY	2732	TCAACGATGAATGGTATTCACAGTAAGACAAAGTGTTCGCCAGAAAATTCCACAAAGCCCTT	2793
Db	2873	TTATGTATGATAGTGTACCAAGTAAAGCCAGATATTGTCTCAGAGCTGCATTAAGGCATTTG	2932
OY	2792	CCCTTTTACGGCTTCCACTTGTAGTATATAGGCAAATCTGCGCCCTTTGTGCCAAAATCTCTG	2851
Db	2933	TGAAGTTACTCTCCCATTTGGAGATATAGGCACTTTTGCCTTGGTGCCAAAACATCTGT	2992
OY	2852	TAAAGGAGAGAGAGCTATCTATGAGCAATTTTGGTGAAAAATATAATGTAAAGCCGG	2911
Db	2993	TGAAGGAGAGAGAGACACAGCCAGACACATTTTACTGAAAAATATCATGTTATCCAGAGG	3052
OY	2912	AGTATCTGAGACAGCATGCACTGTTAGTGAATAATTTGTCTCTCTTACCAAGATATG	2971
Db	3053	AATTCATTAACCAAAATCTTATGGCTACTGTGAATAATTTATTCACCTGTGGCTGAATGTG	3112
OY	2972	TTGTTCATATACAAATTCACCTTTTGGCACATGCCAGATTAATGTCCAAATACAGGATA	3031
Db	3113	TAGTTCCATCATGATGTTTACCCTGCTAGCCCATGATCCAGATTTTACAGATACCAAGTATG	3172
OY	3032	TTGAACAACCTTAAAGATGTTAAAGATGTCTTTGGTTTGTCTGGAATATTAAATGGCTA	3093
Db	3173	TTGATCAGCTTCGTGATATCAAAAGAGTCCATGTTCATGCTTGAAGTTTAAATGACAA	3232
OY	3092	AAATGGAATAATAAGTCACGCTTTTATACAAAAGATGTGAAATATTAAACAAACA	3153
Db	3233	AGAAATGAATAAAATAGCCATCTTTTATGAAGAAAGTGCGAGAAACATCAAGTTAAACA	3292
OY	3152	AAGATGCCCAAGGACACAGATGATCCAAAATAAGATGAATAAACTGTACACTGTGTGATG	3211
Db	3293	GAGATGCCCAAGTCCTCCAGATGAATCCAAAGCAAAATGAATAAACTGTATACATATGTATG	3352
OY	3212	TTGCAATGAATATATCATCTGTCAAAGAGTACTACATACAGTTTGAATCTCTTAAAGAC	3271
Db	3353	TGCGCTCTCTGTGTATTAATATGTAAGTATGCTTTGTGTGCAATGCAGATTCACCAAAAGCAC	3412
OY	3272	CGGATACACAGCTGCTGTTTCTTCACTCAAACTGCAGCAAGATTTTGTAGTAACCAACAAAT	3331
Db	3413	CAGTCCTCCCAATGAATTTTATACACAACTGAAAAGACTTCTGTATACATATGAAGTT	3472
OY	3332	ATCTGCTCTCTGAATGAATCATTTTTCACCTCTCGAANAACGTAAACAACCAATGTTCT	3391
Db	3473	ATATTTTCAAGAAAGACAAAGTACTTCTGTATACAGGAAGCCAAAGCTCTGTGAGATC	3532

QY	3352	TAGGAGCTGTTTAACAGCCACTTTCATCAGACGACCAATCTCAGACCAATCATCAC	3451
Db	3533	TAGTGCGCAGTAAATPAAAGCTTTATCCAGACACGGGAGGAAACCCATGTTTGAAGACACTG	35922
QY	3452	GAATGGAACGCTGAACGAATGCAAGACACGACCTCAAAATCCAAAGCTTCCTCGAAGAATPA	35111
Db	3593	GCACGTGAACCTGGAAGCAATATTTAATGTAAATTTCAGACCTGAACCTTCAACCGGAATTC	36522
QY	3512	AGGGGAGCTTGATAGTCTGAAATGATCAGATCAGTGAATGAAGA	3557
Db	3653	GATCAAGGGAACAGAGATTCAAGAGCCAGACAGAAACTGGAGTTAGTGA	3698
RESULT 7			
HSU50533		1852 bp	Linear
LOCUS			
DEFINITION	Human BRCA2 region, mRNA sequence	CG008.	
ACCESSION	U50533		
VERSION	U50533.1	GI:1531605	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1852)		
AUTHORS	Couch,F.J., Rommens,J.M., Neuhausen,S.L., Bannanger,C., Dumont,M., Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L., Farid,L., Frye,C., Hattier,T., Jannecki,T., Jiang,P., Kehrner,R., Leblanc,J.-F., McArthur-Morrison,J., McSweeney,D., Miki,Y., Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M., Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T., Tranchant,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H., Labrie,F., Stolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L., Tavtigian,S.V. and Simard,J.		
TITLE	Generation of an integrated transcription map of the BRCA2 region on chromosome 13q12-q13		
JOURNAL	Genomics 36 (1), 86-99 (1996)		
MEDLINE	96411650		
PUBMED	8812419		
REFERENCE	2 (bases 1 to 1852)		
AUTHORS	Simard,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular Endocrinology, CHU Research Center, 2705, Boulevard Laurier, Quebec City, Quebec G1V 4G2, Canada		
FEATURES	Location/Qualifiers		
SOURCE	1..1852		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="13"		
	/map="13q12-q13"		
	/note="CG008; DSG numbers: D13S171 and D13S310"		
BASE COUNT	609 a 348 c 379 g 505 t 11 others		
ORIGIN			
Query Match	42.0%: Pred. 1754.4; DB 9; Length 1852;		
Best Local Similarity	98.0%: Pred. No. 0;		
Matches 1817; Conservative	10; Mismatches 21; Indels 6; Gaps 5;		
QY	1064	TTAAAGTAGAGTCACATGACCCCTGAGAAAGTATTAGACATGATTTATGTGCAATAG	11233
Db	1	TTAAAGTAGAGTCACATGACCCCTGAGAAAGTATTAGACATGATTTATGTGCAATAG	60
QY	1124	TTACAGCTGCTAAAGAGATATTTCTTGTCATATGATCACTTACTTAATTTTGTGAGAG	11833
Db	61	TTACAGCTGCTAAAGAGATATTTCTTGTCATATGATCACTTACTTAATTTTGTGAGAG	120
QY	1184	AGAGAACCTTGAACAAAGATGAGAGAGAGACCAAGAAAGCATATGGAGACTTGGCCAAA	1243
Db	121	AGAGAACCTTGAACAAAGATGAGAGAGAGACCAAGAAAGCATATATGGAGACTTGGCCAAA	180
QY	1244	TTTATTAAGAAATATGCTTTACAGTCAGACGCTGGAAAAAGATGCTGCAAAACGATPACAT	13033

Db	181	TTTATTAAGAAATATGCTTTACAGTCAGAGCGTGGAAAAGATGCTGCACAAACAGATAGCAT	240
QY	1304	GGATTCAAAGACAAATTGCTACATATATATATCAAAATAGTATGGATGAGTACTTGG	1363
Db	241	GGATTCAAAGACAAATTGCTACATATATATATCAAAATAGTATGGATGAGTACTTGG	300
QY	1364	TTGAAAGGATCTTGGCTCAATACATAGTGGTCTCCACAAATTTGAAACTCACAACGGATGA	1423
Db	301	TTGAAAGGATCTTGGCTCAATACATAGTGGTCTCCACAAATTTGAAACTCACAACGGATGA	360
QY	1424	AATGCTTATATATCTTGATGTCACACACTGAGATTAAATGCTGTGAAACATATGAATGAA	1483
Db	361	AATGCTTATATATCTTGATGTCACACACTGAGATTAAATGCTGTGAAACATATGAATGAA	420
QY	1484	TGTGGAATATGCCAAATATGCTGCCGACATCAAGTAAGAGATTGGCTTGACTTGATTAAGC	1543
Db	421	TGTGGAATATGCCAAATATGCTGCCGACATCAAGTAAGAGATTGGCTTGACTTGATTAAGC	480
QY	1544	AAACCAAAACAGATCCGACGTGCACAGGCCATATTTTCAAAGTGATGGTATTTACACAAA	1603
Db	481	AAACCAAAACAGATCCGACGTGCACAGGCCATATTTTCAAAGTGATGGTATTTACACAAA	540
QY	1604	ATTTACCTGATCTCGTAAAGCCTCAGATTTTCATGTAAGAAATTCACACAGSGTGTAGAG	1666
Db	541	ATTTACCTGATCTCGTAAAGCCTCAGATTTTCATGTAAGAAATTCACACAGSGTGTAGAG	600
QY	1664	ATGATGACAAATATGAAGACAGTTAGAGTACTTGTTAGTCCACACATGCTCTGCACAGC	1723
Db	601	ATGATGACAAATATGAAGACAGTTAGAGTACTTGTTAGTCCACACATGCTCTGCACAGC	660
QY	1724	AGCGTGAAGGTGGTGGCTGCAAAATATCAAGAAAGTTGGGCAACCCCAACAGCCTACAA	1783
Db	661	AGCGTGAAGGTGGTGGCTGCAAAATATCAAGAAAGTTGGGCAACCCCAACAGCCTACAA	720
QY	1784	ATCCTTCTCTGGAATATGATCAAGTTCTCTTGGAGAGATATAGCACCTGTGCACATAGATA	1843
Db	721	ATCCTTCTCTGGAATATGATCAAGTTCTCTTGGAGAGATATAGCACCTGTGCACATAGATA	780
QY	1844	CCGAATCTATCAGTCTCTTATTTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATG	1903
Db	781	CCGAATCTATCAGTCTCTTATTTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATG	840
QY	1904	ATGAAGATGAGSGTGTCCCAACTGATCAAGCCATCAGACAGCGTCTTAACGTCTTAAG	1966
Db	841	ATGAAGATGAGSGTGTCCCAACTGATCAAGCCATCAGACAGCGTCTTAACGTCTTAAG	900
QY	1964	TACTCTCATTTACACATCCATCTCATTTCAATCTGCGAAACATTTAAATCATATACGG	2023
Db	901	TACTCTCATTTACACATCCATCTCATTTCAATCTGCGAAACATTTAAATCATATACGG	960
QY	2024	CTTGTCTGAAAATGAGATGATGAAAAAGTAGCAGAAAGTGCACCTACCAATTTTCCAAAACA	2083
Db	961	CTTGTCTGAAAATGAGATGATGAAAAAGTAGCAGAAAGTGCACCTACCAATTTTCCAAAACA	1020
QY	2084	CAGGAAGCAAAATGSAAGAGATTTTCCACACATCAGATCAAGCCTTCTCTGTTTAC	2143
Db	1021	CAGGAAGCAAAATGSAAGAGATTTTCCACACATCAGATCAGATCAAGCCTTCTCTGTTTAC	1080
QY	2144	ATTCACAAATCTAAAAAAGAGCCCCCGG -TCAAGCCAATATGTCATTCATTTGATTCAT	2207
Db	1081	ATTCACAAATCTAAAAAAGAGCCCCCGG -TCAAGCCAATATGTCATTCATTTGATTCAT	1140
QY	2203	GCGATATATTTTCTAGTAAAGACAGCAGTTTGCACAGATATTTGAGCCCTGCGATTAAGGC	2266
Db	1141	GCGATATATTTTCTAGTAAAGACATCAAGTTTGCACAGATATTTGAGCCCTGCGATTAAGGC	1200
QY	2263	CTAGATCCAAAGCACTGGAACATCTCTAACACCATGTTACTATTTGGTATATTTGCT	2322
Db	1201	CTAGATCCAAAGCACTGGAACATCTCTAACACCATGTTGGTATTTGGTATATTTGCT	1266
QY	2323	CTCCTTGACCTGATCAATTTGCTGCTCTTTGGAAATCTTGGGTAGCTACTTTCATTTGTG	2382

Db	1261	CTCCTTGACCTGATNCATTTGGTGCCTCTTGAAATCTTTGGTAGCTACTTCATTGyg	1320
OY	2383	AAAGATCTTCTCATGATGATGCGCTTCCAGGGAAGAAAGACACTAACTTTGGGTTCCA	2442
Db	1321	AAAGATCTTCTCATGATGATGCGCTTCCAGGGAAGAAAGACACTAACTTTGGGTTCCA	1380
OY	2443	GATGAGAAGATATCTCCTGAGACAAATGTCAAAATTCAGCTATTAAATGATGGTTGCA	2502
Db	1381	GATGAGAAGATATCTCCTGAGACAAATGTCAAAATTCAGCTATTAAATGATGGTTGCA	1440
OY	2503	TGCGTACTTGGAATGAAAAATTAACAGATTAATCAGAACTTCTACCTTAAG - ATTGCT	2561
Db	1441	TGCGTACTTGCGCATGTAAATTAAATCAGAGTATACAGAGATCTTCTCCTTAAGSATTTGCT	1500
OY	2562	AACAAACAATATTTGCATAGTAGTGAAGACTTGACAGAACAGGGGAAAAATTAGTAACAGA	2621
Db	1501	AACAAACAATATTCGATCGTGTATGGGGGA - TTGGMAGAACAGGGGAAA - TTAGTAACAGA	1557
OY	2622	TATGTCACGCTGAGACTTGTGCTGTGGAGTGCATTTGTGAAGCTGGCACAAAGACCTG	2681
Db	1588	TATGTCACGCTGAGACTTGTGCTGTGGAGTGCATTTGTGAAGCTGGCACAAAGACCTG	1617
OY	2682	TTACCAAGAAATCAATACATATAGAAACAATATACAGCTATGTGCATTTAGCTATCAACAGATGA	2741
Db	1618	TTACCAAGAAATCAATACATATAGAAACAATATACAGCTATGTGCATTTAGCTATCAACAGATGA	1677
OY	2742	ATGCTATCAAGTAGACACAAGTGTTT - GCCCAGAAACTTCACAAAGGCCCTTCCCGTTTAC	2800
Db	1678	ATGCTATCAAGTAGACACAAGTGTTTGGGCCAGAAACTTCGCAAAAGGCCCTTCCCGTTTGC	1737
OY	2801	GGCTTCCACTTGAGTATATGGCAATCTGTGCGCTTTGTGCAAAAGATCTGTAAGAGAGA	2860
Db	1738	GGCTTCCACTTGAGAAATATGSCAATCTGTGCGCTTTGTGCAAAAGATCTGTAAGAGAGA	1797
OY	2861	GAAAGAGCTCATGCTAGGCAATGTTTGTGGAAGAAATATATAATCTAAAGCGGGAGT	2914
Db	1798	GAAAGAGCTCATGCTAGGCAATGTTTGTGGAAGAAATATATAATCTAAAGCGGGAGT	1851

RESULT #				
LOCUS	BC039256			
DEFINITION	BC039256	4252 bp	mRNA	linear PRI_06-NOV-2002
	Homo sapiens,	Similar to androgen-induced prostate proliferative shutoff associated protein, clone MGC:33757 IMAGE:5295109,	mRNA,	
	complete cds.			
ACCESSION	BC039256			
VERSION	BC039256.1	GI:2465778		
KEYWORDS	MGC.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 4252)			
AUTHORS	Strausberg,R.			
TITLE	Direct Submission			
JOURNAL	Submitted (01-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
	USA			
REMARK	NIH-MGC Project URL:	http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk			

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNLD)
DNA Sequencing by: Sequencing Group at the Stanford Human
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shpc.stanford.edu>
Contact: (Dickson, Mark) mcdelpa@llnld.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMLU at: <http://image.lmlu.gov>
Series: IRAC Plate: 48 Row: k Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7705287.

FEATURES

source

1..4252
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:33757 IMAGE:5295109"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="vector: pBluescript"
71..1660
/codon_start=1
/product="Similar to androgen-induced prostate proliferative shutoff associated protein"
/protein_id="AAH39256.1"
/db_xref="gi:24657779"

CDS

BASE COUNT

1407 a 664 c 792 g 1389 t

ORIGIN

1 ATGGCTCATTTAAAGACTAGACCAATGATGAAATTTACATATCCGCTGGGCTCAAG 60
|||||
71 ATGGCTCATTTAAAGACTAGACCAATGATGAAATTTACATATCCGCTGGGCTCAAG 130
|||||
61 GAAATATCAGATTAATAATATCTAAAGAGAGATGAGACGATTAAGATGTTGAAA 120
|||||
131 GAAATATCAGATTAATAATATCTAAAGAGAGATGAGACGATTAAGATGTTGAAA 190
|||||
121 ACTTTATGATATGAGACGACGACTCTGAAGAAGAAAAGAGCTTTATTTAAACCTAGCT 180
|||||
191 ACTTTATGATATGAGACGACGACTCTGAAGAAGAAAAGAGCTTTATTTAAACCTAGCT 250
|||||
181 TTACATCTTGGCTCAGATTTTCTCTCAAGCATCTGCTTAAGANGTTCGCTTACTGCTA 240
|||||
251 TTACATCTTGGCTCAGATTTTCTCTCAAGCATCTGCTTAAGANGTTCGCTTACTGCTA 310
|||||
241 GCGTGTGCTTGGCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 300
|||||
311 GCGTGTGCTTGGCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 370
|||||
301 GATTAACATAAGGATATATTTATGTTTAAACAAGACAGTTGAAGGGGCTAGAGATACA 360
|||||
371 GATTAACATAAGGATATATTTATGTTTAAACAAGACAGTTGAAGGGGCTAGAGATACA 430
|||||
361 AAGAGCCACAAATTCATAGGATTTTATTTACTTGAGACATGCTTGGTGAAGCA 420
|||||
431 AAGAGCCACAAATTCATAGGATTTTATTTACTTGAGAACATTTGCTTGGTGAAGCA 490
|||||
421 TATTAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTACCCAGCTATACAGAAC 480
|||||
491 TATTAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTACCCAGCTATACAGAAC 550
|||||
481 TTATTTTCAGTTATTAACATATGCGCACATCAGAAAGTCCATATGACATGATAGACCTT 540
|||||
551 TTATTTTCAGTTATTAACATATGCGCACATCAGAAAGTCCATATGACATGATAGACCTT 610

Query Match 35.1%; Score 1464.6; DB 9; Length 4252;
Best Local Similarity 95.5%; Pred. No. 2.3e-291;

Matches 1518; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

1 ATGGCTCATTTAAAGACTAGACCAATGATGAAATTTACATATCCGCTGGGCTCAAG 60
|||||
71 ATGGCTCATTTAAAGACTAGACCAATGATGAAATTTACATATCCGCTGGGCTCAAG 130
|||||
61 GAAATATCAGATTAATAATATCTAAAGAGAGATGAGACGATTAAGATGTTGAAA 120
|||||
131 GAAATATCAGATTAATAATATCTAAAGAGAGATGAGACGATTAAGATGTTGAAA 190
|||||
121 ACTTTATGATATGAGACGACGACTCTGAAGAAGAAAAGAGCTTTATTTAAACCTAGCT 180
|||||
191 ACTTTATGATATGAGACGACGACTCTGAAGAAGAAAAGAGCTTTATTTAAACCTAGCT 250
|||||
181 TTACATCTTGGCTCAGATTTTCTCTCAAGCATCTGCTTAAGANGTTCGCTTACTGCTA 240
|||||
251 TTACATCTTGGCTCAGATTTTCTCTCAAGCATCTGCTTAAGANGTTCGCTTACTGCTA 310
|||||
241 GCGTGTGCTTGGCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 300
|||||
311 GCGTGTGCTTGGCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 370
|||||
301 GATTAACATAAGGATATATTTATGTTTAAACAAGACAGTTGAAGGGGCTAGAGATACA 360
|||||
371 GATTAACATAAGGATATATTTATGTTTAAACAAGACAGTTGAAGGGGCTAGAGATACA 430
|||||
361 AAGAGCCACAAATTCATAGGATTTTATTTACTTGAGACATGCTTGGTGAAGCA 420
|||||
431 AAGAGCCACAAATTCATAGGATTTTATTTACTTGAGAACATTTGCTTGGTGAAGCA 490
|||||
421 TATTAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTACCCAGCTATACAGAAC 480
|||||
491 TATTAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTACCCAGCTATACAGAAC 550
|||||
481 TTATTTTCAGTTATTAACATATGCGCACATCAGAAAGTCCATATGACATGATAGACCTT 540
|||||
551 TTATTTTCAGTTATTAACATATGCGCACATCAGAAAGTCCATATGACATGATAGACCTT 610

541 ATGACCTCTATTTATTTGGAAGGTATACAGTGTCTCAGAGACTTTGGATACGTTTAA 600
|||||
611 ATGACCTCTATTTATTTGGAAGGTATACAGTGTCTCAGAGACTTTGGATACGTTTAA 670
|||||
601 GTAATTCGATACCGCTGCATATGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
|||||
671 GTAATTCGATACCGCTGCATATGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 730
|||||
661 TTACTGAAGAGACAGCTCAAGCTATATAGGACATATATTTACACTTTTATATCAGGTT 720
|||||
721 TTACTGAAGAGACAGCTCAAGCTATATAGGACATATATTTACACTTTTATATCAGGTT 790
|||||
721 CTGATGCTTGGGAAACATCTATCAGCGATTTGTCAGAGCATGCTTTGACTTAATTTTG 780
|||||
791 CTGATGCTTGGGAAACATCTATCAGCGATTTGTCAGAGCATGCTTTGACTTAATTTTG 850
|||||
781 GAGCTCTCAATATTTGATATGATATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
|||||
851 GAGCTCTCAATATTTGATATGATATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
|||||
841 TTAAGAGCAATGATATGAGAGAGCGCTTACAGTTGTTAAACTACTGGCAAAATGTTT 900
|||||
911 TTAAGAGCAATGATATGAGAGAGCGCTTACAGTTGTTAAACTACTGGCAAAATGTTT 970
|||||
901 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAAACAGCCACTTTGGCAGTCTACTGGGC 960
|||||
971 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAAACAGCCACTTTGGCAGTCTACTGGGC 1030
|||||
961 AGGTTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 1020
|||||
1031 AGGTTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 1090
|||||
1021 CTCATGAACCATCTGATTTAGCAAAAGACTTAACAGAGATCTTAAAGTGAAGTGCATAT 1080
|||||
1091 CTCATGAACCATCTGATTTAGCAAAAGACTTAACAGAGATCTTAAAGTGAAGTGCATAT 1150
|||||
1081 GACCTGAGGAAGCATATGACATGATATGATATGATATGATATGATATGATATGATATG 1140
|||||
1151 GACCTGAGGAAGCATATGACATGATATGATATGATATGATATGATATGATATGATATG 1210
|||||
1141 GATATCTTCTGTCATGATATGATATGATATGATATGATATGATATGATATGATATGATAT 1200
|||||
1211 GATATCTTCTGTCATGATATGATATGATATGATATGATATGATATGATATGATATGATAT 1270
|||||
1201 CGATGAGAGATGACGCAAAAGAGCCATGATGAGGACTTGGCCAAATTTATTAAGAAATATGCT 1260
|||||
1271 CGATGAGAGATGACGCAAAAGAGCCATGATGAGGACTTGGCCAAATTTATTAAGAAATATGCT 1330
|||||
1261 TTACAGTCAGAGCATGGAAGAAAGATGTCGCAAAACAGATAGAGATGATCAAGACAAATG 1320
|||||
1331 TTACAGTCAGAGCATGGAAGAAAGATGTCGCAAAACAGATAGAGATGATCAAGACAAATG 1390
|||||
1321 CTACATATATATTTATCAAAATAGTATGATATGATATGATATGATATGATATGATATGAT 1380
|||||
1391 CTACATATATATTTATCAAAATAGTATGATATGATATGATATGATATGATATGATATGAT 1450
|||||
1381 CAATACATGTTCTCTCACATTTTAAACAATCAACAAGAGATGAAGTCTTATATTTACTTG 1440
|||||
1451 CAATACATGTTCTCTCACATTTTAAACAATCAACAAGAGATGAAGTCTTATATTTACTTG 1510
|||||
1441 TATGCGCACATGGAATTTAAAGTGTGATGAACCATTTGAATGTAAGTGAAGTGAAGT 1500
|||||
1511 TATGCGCACATGGAATTTAAAGTGTGATGAACCATTTGAATGTAAGTGAAGTGAAGT 1568
|||||
1501 CTGCTCCGACATCAAGTAAAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
|||||
1569 TTCAATCTCTGTCATATATGATATGATATGATATGATATGATATGATATGATATGATATG 1628
|||||
1561 AGTGTCAAGGCCATATTTTCAAAAGTAT 1589
|||||
1629 AATGTAATCTATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 1657
|||||

RESULT 9
 LOCUS AB014548 5177 bp mRNA linear PRI 06-FEB-1999
 ACCESSION AB014548
 VERSION AB014548.1 GI:3327109
 KEYWORDS
 SOURCE ORGANISM Homo sapiens (human)
 REFERENCE 1 (sites)
 AUTHORS Ishikawa, K., Nagase, T., Suyama, M., Miyajima, N., Tanaka, A., Kohana, H., Nomura, N. and Ohara, O.
 TITLE Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
 JOURNAL DNA Res. 5 (3), 169-176 (1998)
 MEDLINE 9743880
 PUBMED 9734811
 REFERENCE 2 (bases 1 to 5177)
 AUTHORS Ohara, O., Suyama, M., Nagase, T. and Ishikawa, K.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: cdnahnfo@kazusa.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)
 FEATURES
 Source Location/Qualifiers
 1..5177
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="H103994"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pluescriptII SK plus"
 /dev_stage="adult"
 1..5177
 /gene="K1AA0648"
 <1..2558
 /gene="K1AA0648"
 /codon_start=3
 /product="K1AA0648 protein"
 /protein_id="BAA31623.1"
 /db_xref="GI:3327110"
 /translation="IYIYVSLIDPNVAKALNEMKCOMLRSHVRELLIDHKOPTSE
 NCSAMPGRKIMIAKINLPDGRKADPQKAFKFNQYICDDEKRLQDLELLISPTSCQADDE
 CVRIARKLANPKOPTNPFLEWAFLEIRIAPIHIDSALVALKLMKKSIEGTADDE
 EEGYSPTAIRNSGLELKVLSFTHTPSHSKEYESLQICLQEMDDKVAEAAIQIFR
 TGHLEINDLPQIRSTLIPILHQAOKARPGPHQAQAVHCHIAIFNKEVQLQIFEPILS
 RSLNADVEQLITPLVSLGHSIMLAPQFASPMKSVYANFVKDLLMDRSTGEKNGK
 LMSDEEVSPEVLAKVOAKILVEMLLGMKNKNSKANSSTIRLSAMVSGDLTEOKR
 RISDSMSRLRLAGSAIMKIAQPCYHEITTPPOLCAIVINDECYOVQITFAOKIT
 HKALVKKLLPLEYNAIFRALCAKDPYKERRARARQCLKNISIRREYIKONNSHAFKKLL
 SLLEIYVPTIHLIHLADPDFTRSDVDQDLIDIECLMFMLEVLTMKENNSHAFMKK
 MAENIKILTRQAQSDSEKSTNEKLTVCDDVALCVINSSALCNADSPKDPVLPKMFETQ
 PEKPFQCNKDSYISIESEFVLLLTGKPRGAVGVANIKPLSATGRKPYVSTGETGSGNI
 NVNSELNSTGNSREOSEAETGVSENEENPYRIISVTPVKNIIDPKKINSDQK
 TQGNISLSDRCKRRTVTAAGAENIQOKTDEKVEDSGSPAPKSPRRGGRPKSSQGNATKK
 MDDLAKPIPKRKRRAAVQGESPGLEAGNAKAPKPLQDLAKAAFAEROIDLQR"
 BASE COUNT 1690 a 931 c 1037 g 1519 t
 ORIGIN
 Query Match 25.0%; Score 1041.4; DB 9; Length 5177;
 Best Local Similarity 68.0%; Pred. No. 3.5e-204;
 Matches 1450; Conservative 0; Mismatches 681; Indels 0; Gaps 0;
 Db 1 GCTATATTACTTATATCTGATGGCACACTGGATTTAAATGCTGTGAAGACATGTAATGAT 1486
 1487 GGAATATGCAAAATCTGCTCCGACATCAAGTAAGAGATTTCCTTGACTTGATTAAGCAAC 1546

D	b	61	GGAGGTGCAAGACATGCTTGGAGGCATGTACCGCAACTATTGGATTGTGCACAGAGC	120
O	y	1547	CCAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTACAGAAATT	1606
D	b	121	CTACATCAGAGGCTAACTGTTCTCCATGTTTGGAAACATGATGCCATATGCAAAAGATT	180
O	y	1607	TACCTGATCCGCTAAGGCTCAGGATTTTCATGAAGAAATTACACAGAGTGTGGAAGATG	1666
D	b	181	TGCTGTACCCCGGAAAGCACACAAATTTTGTGAAGAAATTTAACACAGGTTCTCGGCATG	240
O	y	1667	ATGAGAAATATAGAAAGCAGTTATAGACTTCTGTTAGTCCAAACATGCTCTCGCAAGCAG	1726
D	b	241	ATGAGAAACTCGGCTCAGTTGAGTAGTTATTAATTACGCCAACCTGTTCTTGCAACAG	300
O	y	1727	CTGAAGTGTGTGCGGTGAATTAACTAAGAGTTGGGCACCCCAACAGCTTACAANTC	1786
D	b	301	CAGATATTTTGTGAGAGAAATATACCCGGAACTTGCAAACTTAAGCAACCAACAATC	360
O	y	1787	CTTTCGGAATGATCAAGTTCTCTCTGGAGAGATAGACACCTGTGACATTAATACCG	1846
D	b	361	CTTTCTAGAAATGGTCAAAATTTCTGTTGGAAACATGCAACCTGTGCACATTGATTCAG	420
O	y	1847	AATCTATCAGTCTCTTATTTAAACAGTGAACAATCAATATGATGAGAAACAGACAGATGATG	1906
D	b	421	AAGCATATAGTGCACCTAGTGAAATTTGATGAATAGTCAATAGAGGGGACACAGATGATG	480
O	y	1907	AAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGTCTTGAACCTGTTAAGGTAC	1966
D	b	481	AAGAGGAGGGGTGAAGTTCACATATACAGCTATCCGTTACAGCACTTGAACCTTTAAGGTTC	540
O	y	1967	TCTCATTTACACATCCCATCTCATTTCTATTGTGCTGAACATTTTGATCATTTACTGGCTT	2026
D	b	541	TGTTCTTTTACCATCTCCACCTCGTTCCACTGTGAGAGCAATTAAGATCTCTTTGTTACAGT	600
O	y	2027	GTCGAAATATGATGATGAAAGATAGACAGAGCTGCACTCAATATTTTCAAAACACAG	2086
D	b	601	GCCTAAGATATGAGATATGACAGAGTACGAGAGCTGCTATTCCAATTTTGAAGAAATACG	660
O	y	2087	GAAACAAATTTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCTGTTTACATC	2146
D	b	661	GTCACAAAATATGAAACACAGACTTCCCGCAGATGACATGACACTTAAATCCCATTTTACATC	720
O	y	2147	ACAATATCAAAAAAGAGCCCCCGCTCAAGCCAATATGCATTCATTGTATTCATGCGA	2206
D	b	721	AAAAAGCAAAAGGGGTACTCTCACACCAAGCAAAACAGCGTGCACCTGATATACGCCA	780
O	y	2207	TATTTTCTAGTAAAGAGACCCAGTTTGGACAGATATTTGACCCTCTGCATPAAGAGCTTAG	2266
D	b	781	TATTCACAAATTAAGAAAGTCCAGCTTGCACAGATTTTGAACCACTCAGTAGAGTCTGA	840
O	y	2267	ATCCAGCAACCTGGAACATCTCATTAACACCATTTGTTACTAATTTGGTCATATGTCTCC	2326
D	b	841	ATGCTGATGTCCCAAGAACAACTTATTAATCTCATTAAGTTTCAATTGGGCCACATTTCTATGT	900
O	y	2327	TTGCACCTGATCAATTTGCTGCTGCTCTCTTGGAAATCTTGGGTGCTACTTTCAATTTGTGAAG	2386
D	b	901	TAGACACAGATCAGATTTGTCTTCCCAATGAATAATCTGTAGTAGCAAAATTTTATTTGTGAAG	960
O	y	2387	ATCTTCATCAATGATGAGCTTCCAGAGGAAAGAACAACTTAACCTTTGGGTTGCCAGATG	2446
D	b	961	ATCTGCTATATATATCACAGTCAACAGGTGAAAGAAATGGAACACTGTGGTCTCCAGATG	1020
O	y	2447	AAGAAGTATCTCTGAGACAAATGGTCAAAATTCAGGCTATTAAATGATGATGGTTCGATGCG	2506
D	b	1021	AAGAGGTTTCCCTGGAAGTACTAGCAAAAGGTACAGGCAATTAACCTCTGTAAAGTGGC	1080
O	y	2507	TACTTTGCAATCAAAATATATACAGTAAATACGAACTTCTACTTAAAGATTGCTAACAA	2566
D	b	1081	TGTTGGGTATTAATAAACACAGCTCTAAATCTGCCAATCAACCTTCGGTTATTATTAACG	1140
O	y	2567	CAATATTGCATAGTAGTAGGAGCTTGAACAGAACAGGGGAAATTAAGTAAACACATATGT	2626

Db 1141 CGATGTGGTAGTGAGGGGTGACCTGACAGCAAAAGAGATCAATATCTGATATGT 1200
 2627 CAGCTGTGAGACTTGTCTGGAGTGCATATGTGAAGCTGCACAAAGAACCTGTACC 2686
 Db 1201 CTCGGCTGGATTAGTCTGCTGGTAGTGCCATATGATGAGCTTCACGAGAACCTGTGTACC 1260
 2687 ATGAAATCATCATATTAGAACAAATATACAGCTATGTGCTATTACTATCAAGTAATGCT 2746
 Db 1261 ATGAAATTTATACCCAGAAAGATTTCACCTCTGTGCACCTTTTATATATGATGATGCT 1320
 2747 ATCAAGTAAGCAAGTGTGGCCCAAGAACTTCACAAAGCCCTTCCGTTACGGCTTC 2806
 Db 1321 ACCAATGAGGAGATATTTGCTCAGAACCTCATAGAGCATGTGGAAGTTACTGCTCC 1380
 2807 CACTGAGTATATGCAATCTGCTGCTTGTGCAAAAGATCTGTAAGAGAGAAAG 2866
 Db 1381 CATTGGAGTATATGGGATCTTGTGCTGTGTCGCAAAATCTCTGGAAGGAGAGAAAG 1440
 2867 CTCATGCTAAGGCAATGTTTGTGAAATAATATATGTAAGCGGAGATCTGGAAGCAGC 2926
 Db 1441 CACAGCAGCAACAATGTTTACTGAAATAATATCAGATATACGAGGGAATCATTAAGCAGA 1500
 2927 ATGCAAGCTGTAGTAAAAATATATGCTCTCTCTACAGAGATATGTTCCATATACAA 2986
 Db 1501 ATCCATGGCTACTGAGAAATATATATCATCTGTGCTGGAATATGATGTTCCATCATGA 1560
 2987 TTCACCTTTTGGCAGATGACCAAGATATATGTAAGTACAGATATGTAACAACCTTAAG 3046
 Db 1561 TTCACCTGCTAGCCCATGATTCAGATTTTACAAAGATCACAGATGTGTGACGCTTCGTG 1620
 3047 ATGTTAAGAATATGCTTTGTTTGTGGAATATTAATGCTAAAAATGAAATATACA 3106
 Db 1621 ATATCAAGAGTGCCTATGCTTATGCTGGAAGTTTATATGACAAAGATGAAACAAATA 1680
 3107 GTACAGCTTTTATGCAAGATGTGTAAGAAATATTAACAACAAAGATGCCCAAGCAGC 3166
 Db 1681 GCCAATGCTTTATGTAAGAGATGCGCAAGAACATCAAGTTTACCAAGATGCCAGTCTC 1740
 3167 CAGATGATGCAAAAAATGATGAAACATGATGATGATGATGATGATGATGATGATGATGAT 3226
 Db 1741 CAGATGATCCAAAGCAAAATGAAACATGATGATGATGATGATGATGATGATGATGATGAT 1800
 3227 TCATGTCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3286
 Db 1801 TAAATAGTAAAGTCTGCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGAT 1860
 3287 GTTCTTTCATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3346
 Db 1861 AATTTTTCACAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 3347 TGAATTCATTTTTCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3406
 Db 1921 CAAGAGTACTTCTGTTAAAGAGAAAGCCAGAGCTGCTGAGTACTAGTGTAGTAATAA 1980
 3407 AGCCATTTTCATGACGACGACGACGACGATCTCAGACCAAAATCATCAGAAATGTAAG 3466
 Db 1981 AGCCTTTATCAGCAAGGAGGAGAAACCTTATGTTAGAAAGCAGCTGAGACTGAGAG 2040
 3467 GCAATGCAAGCAGCAGCTCAAAATCAAGCTCTCTGGAAGAAATTAAGGGAGAGCTTGATA 3526
 Db 2041 GCAATATTTAATGTAATTCAGAGCTGAAACCTTTCAACGGAAATGATCAACGAGAAACAGA 2100
 3527 GTTCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3586
 Db 2101 GTTCAGAGGACGACGAAACTGGAGTTAGTGA 2131

RESULT 10
 BC021408 1976 bp mRNA linear ROD 16-APR-2003
 LOCUS Mus musculus RIKEN cDNA 9030416H16 gene, mRNA (cDNA clone MGC:29439
 DEFINITION IMAGE:3964501), complete cds.
 ACCESSION BC021408

VERSION BC021408.1 GI:18204099
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 1976)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheet, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullan, S.J., Bosak, S.A., McSwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shewchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22368257
 12477932
 2 (bases 1 to 1976)
 Strausberg, R.
 Direct Submission
 Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://imgc.nci.nih.gov>
 REMARK
 COMMENT
 Contact: MGC help desk
 Email: cgabbs@remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LUNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legspt, R., Maduro, O.L., Mastello, C., Maskerli, B., Mastrian, S.D., McKillop, J.C., McDowell, J., Pearson, R., Stantip, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LUNL at: <http://image.llnl.gov>
 Series: IRAX plate: 39 Row: b Column: 24
 This clone was selected for full length sequencing because it passed the following selection criteria: GenomScan gene prediction.
 FEATURES
 source
 location/Qualifiers
 1..1976
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:29439 IMAGE:3964501"
 /tissue_type="mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam1"

Query Match	Best Local Similarity	Matches 1297	Conservative	Score 937.6	DB 10	Length 1976
1468	AAAGCATTGAATGAAATGTGAAATGTGCAAAATCTGCTCCGACATCAAGTAAAGATTG	1527	1	1976	1	1976
25	AAAGCTCATAGAAATGTGAAATGTGCAAAATCTGCTCCGACATCAAGTAAAGATTG	1527	1	1976	1	1976
1528	CTTGACTTGATTAAGCAACCCAAAACAGATGCCAGTGCACAGCCATTTTCAAAAGTG	1587	1	1976	1	1976
85	TTGGAGCTTACAAAGCAGCCTTACATCAAAACCGAATCTGTCGACATGTTGGAAACTG	144	1	1976	1	1976
1588	ATGGTATTACAAAGAAATTTACTGATCTGTAAGGCTCGAGATTCATGAGAAATTC	1647	1	1976	1	1976
145	ATGACACATGCAAAAGAAATTTGCTGACCTCGAAGCAAGTATTTGTAAGAAATTT	204	1	1976	1	1976
1648	ACACAGGTGTGAGAGATGATGAGAAATTAAGAAAGCAGTTGAGATTAATTGTAATCA	1707	1	1976	1	1976
205	AACGAGGTCTTGATGATGATGAGAAATGAGTCTGACGTGGAATTAATTAATGACCCA	264	1	1976	1	1976
1708	ACATGCTCTCTGCAAGCAGCTGAGAGTGTGTGCTGAAATTAATTAAGAGTGGCAAC	1767	1	1976	1	1976
265	ACCTGCTCATGCAAGCAGCTGAGAGTGTGTGAGGAGAAATTAAGTCTGCAAAAT	324	1	1976	1	1976
1768	CCCAACAGCCTTACAAATCTCTTCTGGAATGATCAAGTTCTCTGGAGAGATGACA	1827	1	1976	1	1976
325	CTTACAGCAGCAGCAACCCCTTTCTCTGAGATGCTCAATTTCTTTGGAAAGAAATGCT	384	1	1976	1	1976
1828	CCGTGACATAGATACCGAATCTATCATGCTCTTATTAACAAGAGTAACAAATCATATA	1887	1	1976	1	1976
385	CCTGAGCAGATGATGAGAGCAGATAGTCACTGAGTAAATTAAGATGATGATGATGATG	444	1	1976	1	1976
1888	GATGAGACAGCAGATGATGAGAGGAGGTTCCAACTGATCAAGCCATGAGACAGGT	1947	1	1976	1	1976
445	GAAAGGAGAGCGGATGATGAGAGAGGAGGTTCTGAGTCCAGATCAGCCATCCGCTCAGA	504	1	1976	1	1976
1948	CTTGAGCAGCTTAAAGTACTCTCAATTTACATCCATCCATTTTCAATTCGCTGGAACA	2007	1	1976	1	1976
505	CTTGAGCAGCTTAAAGTACTCTCTCTTTCACATCCATCCATTCGCTGCAAGAGCA	564	1	1976	1	1976
2008	TTTGATCATTAAGGCTTCTGCTGTAAGATGATGAGAAAGTAAAGTAAAGTAAAGTAAAGT	2067	1	1976	1	1976
565	TATGAGTCTCTGTTAACAATGCTGTAAGATGAGATGAGATGAGATGAGATGAGATGAGAT	624	1	1976	1	1976
2068	CAAAATTTTCAAAAACAGAGCAAAATTAAGAGATTTTTCACACATCAGATCAGCC	2127	1	1976	1	1976
625	CAAAATTTTCAAAAACAGAGCAAAATTAAGAGATTTTTCACACATCAGATCAGCC	684	1	1976	1	1976

OY	2128	TTGGTCTCTGTTTACATCACAAATGCAAAAAAGGACCCCCCTCTCAAGCCAAATATGCC	2187
Db	685	TTGATCCCATTTTACATCAGAAAAGCCAAAGGGGGGACTCCACACCAAGCAAGCAAGGCT	744
OY	2188	ATTATATTGTAACCATGCGATTTTCTTCTAGTAAAGACCCAGTTTGGACAGATTTTGGAG	2247
Db	745	GTTACATGCAATCCATGCCATCTCTCTCAAAACAGAGAGGTCCACAGCTGGSCACAGATTTTGGAG	804
OY	2248	CCCTCGCATTAAGACCTAGATCCAAAGAACCTGGAAATCTCATACACACTGGTTACT	2307
Db	805	CCACTCAGTAGACTCTGGAATGCTGATGTAACAGACGAACTTAACTCCATTAGTTTCA	864
OY	2308	ATTGGTCATATTGCTCTCTCTTGCACCTGATCAATTTGCTGCTCTTGGAAATCTTGGGTA	2367
Db	865	CTGGGCCACATTTTCCATAGTTAGACACAGATCAGTTTCTCTCCCGATGAATATCTGTAGTG	924
OY	2368	GCTACTTTTCAATGTAAGAAAGATCTCTCATAGTAATATGGGCTTCCAGGGAAAAAGCACACT	2427
Db	925	GCAACCTTATTGTTTAAAGACCTTCTATATGACACACAGGTCAACAGGTGAGAAAGATGCA	984
OY	2428	AAACTTGGTTCAGATGAAGAAAGTATCTTCTGAGACAAATGTCAAATTTACGGCTATT	2487
Db	985	AAATTATGCTTCCAGATGAGAGAGATTCTCTCTGAATCTTACGAAAGATGACAGCAATT	1044
OY	2488	AAATNGATNGTTCAGTGGCTACTTGGAAATGAAATATATCAGTAATATAGGAACCTCT	2547
Db	1045	AAACTCTGGTAAGGTGGCTGTTGGATGAAACCAACAGTCTAATATCTTCCCACTCA	1104
OY	2548	ACCTTAAGATTGCTAACACAAATATTGCATAGTATGAGAGACTTGACAGCAAGGAGAA	2607
Db	1105	ACCTTGCGGTATTATACAGCATCTGTGATGATGAGGGCGACCTGACAGAGCAAAAGAG	1164
OY	2608	ATTAGTAACCAAGTATGTCAAGCTGAGACTTCTCTGGGATGCTATTTGGAAGCTG	2667
Db	1165	ATACGTAAATGTGATGTCTCGCTTGGATGACTCTGTGATGTCATATGAAACCTT	1224
OY	2668	GCACAGAGACCTTTTACCATGAATATCATCATTAAGAAATATACATGACTATGCAATTA	2727
Db	1225	GCTCAGGAACTTGCTACCATGAATATTATACCCAGAACAGTTTCAGCTCTGTCAGCTG	1284
OY	2728	GCTATCAACATGATGCTATCAAGTAAGACAACTGTTTGGCCGAAACTTCACAAAGCC	2787
Db	1285	GTTATTAATGATGATGTGTACCAAGTAAGGACGATTTTGGCCAGAAAGCTTCATTAAGCA	1344
OY	2788	CTTTCGCCGTTACAGGCTCCACTTGAGATATATGCAATCTGTCGCCCTTTGCGCAAAAGAT	2847
Db	1345	CTTGGAAGTTGCTTCTCCCACTGGAGATATATGSCCATCTTCTGTTGTGCGCAAGAC	1404
OY	2848	CTCTGTAAGGAGAGACAGCTCATGCTAGAGCAATGTTTGGTGAATAATATTAATGTAAAG	2907
Db	1405	CTCTGTAAGGAGAGAGAGACAGAGCTCGGAGGTGTTGTTAAAGAACATCAGCATCCGC	1464
OY	2908	CGGAGATATCGAAGACAGCATGACGCTGTAAGTGAATAAATATATGCTCTTACCCAGAG	2967
Db	1465	AGGAGATGATCAACACAGAACCCCATGCGCACTGAGAAATTAATGTACTGCTGCTGAA	1524
OY	2968	TATGTTGTTCCATATACAAATCACCTTTGGCAGATGACCCAGATTAATGTCAAAAGTACAG	3027
Db	1525	TATGTTGTTCCATATGATATCACCTCTACCCATGATCTCTGATTTCCACAGGTCACAA	1584
OY	3028	GATATGTAACAACTTAAAGATGTTAAAGAAATGTTTGGTTTGTCTGGAAATTTATATG	3087
Db	1585	GATGTTGTAACAATCTGATATTAAGAGAGCGCTGTGGTTTATGCTTGAAGTCTTATATG	1644
OY	3088	GCTAAAAATGAAATTAACAGTACAGCGCTTTTATCAGAAAGATGTAGAAATATTAACAA	3147
Db	1645	ACAAAGATGAAGAAACACAGCCATGCTATGAAAAAGATGCGAGGAATATACACTA	1704
OY	3148	ACAAAAGATGCCAAGGACAGATGTCAAAAATGAAATGAAATGTAACGTGTGTGT	3207
Db	1705	ACCGAGAGCGCCAGTCTCCCGATGAATCCCAAGCAATATGAATAAATCTTATACGTTGTGT	1764

OY		988	CGCGTGGAAAGTGGAATTTGCTAGCCATTGTCTCAAGAACCTTCGTATTTGCAAA	1047
Db		1142	AGATTAGAAGTAGTGAATTTGCCAGTCATTTGTTAATGANTACCCAGATTTAGCAG	1201
OY		1048	GACTTAACAGAGTATCTTAAAGTAGAGGTCAACATGACCCTGAGAAAGCATATTTAGCATGAT	1107
Db		1202	GATCTCACAGAAATATTTAAAGGTATGATGCATGATCCAGAAAGCATATTCGTGATGAT	1261
OY		1108	GTTATTTGTGTCAATAGTTACAGCTGCTAAAAAGATATTCTCTCGGTCATGTGACTTGA	1167
Db		1262	GTCATTTGTCTATAATTAACAGCTGCCAAGAGGGAGCTCGGCTTAGTAATTAATGATCAGCTG	1321
OY		1168	CTTATATTTTGTGAGAGAGAGAACATTTAGACAACAGATGAGAGATACCAGAAAGACCATG	1227
Db		1322	CTTGCGCTTTGTGAAGGGAAGAAAGACACTGCGATTAACCGTGGCGAGTAAGAAAAGAAAGCATG	1381
OY		1228	ATGGGACTTGGCCCCAATTTTATAGAATAATGTGCTTTACAGTCAGCAGCTGGAAAAGATGCT	1287
Db		1382	ATGGGTCTGGCGTCAAGTTTTATTAAGAATACTGCTTCATGGTGAAGAGGAAAGAAAGCT	1441
OY		1288	GCAAAACAGATAGATGATGATCAAGACAAATTTGCTACATATATTTATCAAAATAGATAT	1347
Db		1442	GCAAGAAAGACAGCTGATTAAGAACAACTTCGTGATATTTATTCAGAACAGCAT	1501
OY		1348	GATGATGACTACTTGTGTTGAAGCGATCTTTGCTCAATAACATGTTCTCACAATTTAGAA	1407
Db		1502	GACACACAACACTGTTGCTAGAGAAAATCTTTGCTCAGATATCTTGTCGCCACACACTG6AA	1561
OY		1408	ACTACAGAAAGGATGAATAAGTCTTATTAATTAATCTTGATGCCACACACTGATTTAAATGCTGTG	1467
Db		1562	ACAGAAAGAGAAATGAATGCTTATTAATTAATGCTATGATGCTAGTCCAAATGCTGTA	1621
OY		1468	AAAGCATTTGATGAATAGTGGAAATGTCCAATTTGCTGCTCGCAKATCAAGTAAGGATTTG	1527
Db		1622	AAACCTCTCAACGAAATGTGGAAAGTGTGACAACTGCTTGAGACCAATGTACGGAACTA	1681
OY		1528	CTTGACTTGGATTAAGCAACCCAAAAGAGTSCCAGTGTCAAGGCGCATATTTTCAAAGTG	1587
Db		1682	TTCGATTTTGCACAAAGCAAGCCATACATCANAGGCTAACTGTTGTCGATTTGGAAAAGTG	1741
OY		1588	ATGCTTATTTACAGAAATTTACCTGATCTGCTGTTAGGCTCAGGATTTTCANAGAAATTC	1647
Db		1742	ATGACCCATWAGCAAGAAATTTGCTGACCCCAGGAAACACAAAGATTTTGTGAAGAAATTT	1801
OY		1648	ACAAGAGCTTTAGAAAGATGATGAGAAAATTAAGAAAGCAGTTAGAAAGTCTGTAGTCCA	1707
Db		1802	AACAGAGCTTCGCGCGATGATGAGAAAACCTTGCTCAGTTGGAGTATTATTAATTAAGGCCA	1861
OY		1708	ACATGCTTCCTGACAGCAGGCTGAAGGTTGTGTG	1740
Db		1862	ACCTGTTCTTCCAAACACAGCAGATATTTGTGTG	1894
<hr/>				
RESULT 12				
RCO41361				
LOCUS				
DEFINITION				
Homo sapiens, similar to KIAA0848 protein, clone MGC:43818				
IMAGE:5273075, mRNA, complete cds.				
ACCESSION				
BC041361				
VERSION				
BC041361.1 GI:27552793				
KEYWORDS				
MGC.				
SOURCE				
Homo sapiens (human)				
ORGANISM				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE				
1 (bases 1 to 2705)				
AUTHORS				
Strausberg,R.				
TITLE				
Direct Submission				
JOURNAL				
Submitted (16-DEC-2002) National Institutes of Health, Mammalian				
Gene Collection (MGC), Cancer Genomics Office, National Cancer				
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
USA				
REMARK				
NIH-MGC Project URL: http://mgc.nci.nih.gov				

COMMENT

Contact: MGC help desk
Email: gcaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letlicia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Scheil, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES

source

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.nl.gov>
Series: IRAC Plate: 75 Row: c Column: 19.
Location/Qualifiers
1..2705
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:43818 IMAGE:5273075"
/tissue_type="Testis"
/clone_id="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
545..2347
/ccoon_start=1
/product="Similar to KIA0648 protein"
/protein_id="AAH41361.1"
/db_xref="GI:27552794"
/translation="MDPTAOPKPATALCGVVSADGKIAYPGVEKITDITIDEMIKRRKRWYQFKFMDSDSEDEKKOYLPLAHLSSEFLRPNPNDRLVLCAGDIERYAPAPAWSHDKLIKLFILFTIRLOLKLEPTKSPFRNYVILLNLAWSYNCEPLEECCNEIFQLFRTLVSTNNSHNKVKYMHLDMSSIIIMGDDVTOELDSLINLIIPARKNLNQSFRLAKVLKRVTVTEACTIANFNOVLVGRSVDSSEHFVDLIOLFLAIDIHPLVSLVPOLFELKINSNGDERLAVALRLAKLGSDSLATONRPLOCFERENDIHYEVLESVKEFASHCLMNHPDLAKLDLEYLKYSHDPEDAIRHDVLTITTAKRDLALVNDOLLGFVRERTLDKRMVRREKAMGAOLYKCYCLGEAGEAKEVSKTMKDRLHITYONSIDSIDKLILEKIEIPAOYLIVPHNLFEERRKCIYLTXASDPNAVVALNMKKCKÖNNLRSHVRELDDHKOPTSANCSAFGKMLMTIAKMLPDGKAQDFVKKFNQVLGDDEKRSQLELLISPSKSCODADICVVS5FYTLFL"

CDS

BASE COUNT	751 a	589 c	654 g	711 t
ORIGIN				
Query Match	22.3%	Score 932.2;	DB 9;	Length 2705;
Best Local Similarity	71.5%;	Pred. No. 1.3e-181;		
Matches 1225;	Conservative 0;	Mismatches 488;	Indels 0;	Gaps 0;
Dy	28	GATGAATAATTTACATATACCGCTGGGGTCAGAAGAAATATCATATAAATATTCAAGAG	87	
Qy	602	GACGGGAAGATCGCTTACCTCCGGGGGTAAAAGAGATCACCGACAAGATACCAAGCAG	661	
Dy	88	GAGATGGTGAGACGATTAAAGATGGTGTGAAAACTTTTATGATATGAGACAGACTCT	147	
Qy	662	GAGATGATCAAAGCCCTGMAAGATGTAAGTAAACCCTTTATGATATGATGATGACAGCA	721	
Dy	148	GAGAAGAAAAGAGCTTATTATTAACCTAGCTTTACATCTTGCCTAGATTTTTTCTC	207	
Qy	722	GAAATGAAAAAACAGCATATCTCCCACTACCTTGATCTTGATCTGCAATCTTCTC	781	
Dy	208	AAGATCTCTGTTAAAGATGTTCCCTTACCTTAGCTGATGCTGCTGCTGATATTTTACG	267	
Qy	782	AGCAACCCCAATAAAGATGTCCTCTCTGTACACATGTTGTTGGCTGGAATCTTTCGT	841	
Dy	268	ATTATGCTCTGGAAGCCCTTACACATCCCTGATATAACATAAGAGATATTTATGTT	327	
Qy	842	ATCTATGCCCCAGAGCTCCATATATCTTCATATATAACTTAAGACATATTTTGTGTT	901	

QY	328	ATACAGACAGCTTGAAGGGCTGAGATCTCAAGAGCCCAATTCATAGGATTTT	387
Db	902	ATTACGACATTTAAAGGTTTGAGGATACAAAGACTCCAGCTTTAAATGATCTTT	961
QY	388	TATTTACTTGAAACATTCCTTGGGTCAGTCATATAACATATGCTTTGAGTTAGAACAT	447
Db	962	TATTTATTAGAATTTAGCTTGGGTTTAAATCATATATACATCTGCTTTGAAATGGAAAT	1021
QY	448	AGCAATGAATTTTCACCCAGCTCTACAGAACCTTATTTTACGTTTAAACAATGGCCAC	507
Db	1022	TGCATATAAATTTTATTTATTCAGCTTTTAAAGAACTCTCTTCAGATGATCAACAAAGCCAC	1081
QY	508	AATCAGAAAGTCATATGACACATGATGAGCTTTATGAGCTCTATTTATTTGGAAGGTAT	567
Db	1082	AATTAAGAAAGTACAAATGACACATCTGATTTGATGAGTCTTATCATCATATGGAAGGTAT	1141
QY	568	ACAGTGTCTACGAGGCTTTTGGATACGGTTTACTAAATCTGGTACCTGCTCATATGAAAT	627
Db	1142	GGAGTTACTCAGAAATTTATGGACATCTTATTTAACTCATCTCCGACATATGAAAC	1201
QY	628	TTAAACAGCAGCATATGATTTTGGCAAGGCTTTACGAAGAGACAGCATCAGCTATT	687
Db	1202	TTAAATTAACAGTCTTTGACCTTGCAAAAGTGTATTTGAAAGAACGTCACATATTT	1261
QY	688	GAGCCATATATTACACCTTTTATTTAACTCAGTTCGATGCTTTGGAAACATCTATACAGC	747
Db	1262	GAGCATGACATGCTAATTTTTCATATCAATCAATCTCGTGCGTGGGAAGATCATCAGTAGT	1321
QY	748	GATTTGTCAAGACATGCTTTGACTTAATTTTGGAGCTCTACATATTTGATAGTCATTTG	807
Db	1322	GATTTGTCAAGAACATGATTTGATCTGATCTGATCAGAACCTTTTGGTATATGATCCCATTTA	1381
QY	808	CTGCTCTCTGTTTATACCCAGCTTGAATTTAAATTAAGACATGATATAGAGAGAGCG	867
Db	1382	TTATATCCGTCATGCCACAGCTTGATTTCAACTTAAGACATGATAGGAAGAGGCA	1441
QY	868	CTACAGTTGTTAACTACTGCGCAAAAATGTTTGGGCAAGATTCAGAAATGGCTTCT	927
Db	1442	TTAGCTTGTGTCCACTTCAGCTAAATTTGTTGGCTCCAAAGATTTGATTTGGCAACA	1501
QY	928	CAAAACAAAGCACATTTGGCAGTGTACTTGGGCAAGTTTATGATATCCATGTCACATC	987
Db	1502	CAGAAATGCTCTCTTTGGCAATGTTTCTTGAGAGATTTAAATGATTTCAATGTTCTGTG	1561
QY	988	CGCCTGGAATGTGGAATTTGCTACGACATGTCATGAACCAATCCGATTTAGCAAAA	1047
Db	1562	AGATTTAAGAAAGTGAATTTGCCACGATCTGTTTAAATGAATCAACCCAGATTTAGCGAAG	1621
QY	1048	GACTTAACAGATATCTTAAAGTACGTCACATACCTTGAGGAAGCTATTAAGACATGAT	1107
Db	1622	GATCTCACAGAAATTTTAAAGGTAGATCAATATATCCAGAAAGAGCTATTTGTCATGAT	1681
QY	1108	GTTATTTGTCAATAGTTACAGCTGCTTAAAGAGATATTTCTTGTCATATGATCACATTA	1167
Db	1682	GTCATTTGTACTAATAATAACGTCGCCAAGAGGAGCTCGGCTTAGTAAATATGATCAGCTG	1741
QY	1168	CTTATATTTTGTGAGAGAGACACTTATGACAAACGATGAGAGATGACCAAGAAAGCATG	1227
Db	1742	CTTGGCTTTTGAAGGAAAGAACTCTGATTAACGGTGGCGCAGTATGAAAGAAAGCTATG	1801
QY	1228	ATGGGACTTTGCCCAATTTTATTAACAATATCTTTACAGTCAGAGAGCTGGAAAGATGCT	1287
Db	1802	ATGGGTCTGCTCAGCTTTTATTAAGAAATACGTCTTCAATGGTGAAGCAGGAAAGAGAGCT	1861
QY	1288	GCAAAACAGATAGCATGATCAAAAGACAAATTTGCTACATATATATTAACAATAATGATTT	1347
Db	1862	GCAAGAAAGTACAGCTGATTAAGAAACAACTTCTCATATTTATTAATCAAAAGACATTT	1921
QY	1348	GATGATGACACTTGTGGAAGGATCTTGGCTCAATATGATGCTCCCAATATTTAGAA	1407
Db	1922	GACGACAAACGTTGTGAGCAAAAATCTTTGCTCAGATCTTGTGCCCCACACACCTGGAA	1981

QY	1408	ACATCAGACAGGAAATGCTTATATTAACCTGTGACACATGGAAATTTAAATCGTGTG	1467
Db	1982	ACAGAAAGAGAAATGAATTCCTTATATTAATCTTAATATCTTAATGGATCCAAATGCTGTGA	2041
QY	1468	AAAGCATTTGATGAAATGTGGAAATGTCCAAATTTGCTCCGACATCAAGTAAGAAGATTTG	1527
Db	2042	AAAGCTCTCAACGAAATGTGGAACTGTCCAGAACATGCTTCGGAGACCATTGACGGGACTA	2101
QY	1528	CTTGACTTTGATTTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAGTG	1587
Db	2102	TTTGATTTTGGCAACAGCAGCCTACATTCAGAGGCTTAAGTGTTCCTGCCATGTTTGGAAACTG	2161
QY	1588	ATGGTTATTTACAGAAATTTTACTGCTCTGTGTGTAAGCTCAGATTTTCATGAGAAATTC	1647
Db	2162	ATGCACATATACCAAGAAATTTTCCGTACCCCGGGAACACCAAGAAATTTTGTGTGAAGAAATTT	2221
QY	1648	ACACAGGTGTGAAGAATGATGAGAAAAATTAAGAAAGCAGTTAAGAAGTACTTGTATGCCA	1707
Db	2222	AACCAAGGTCTTCGGCGCATGTGAGAAAACTTCGGTCTCAGTTGGAGTATTAATTAAGGCCA	2281
QY	1708	ACATGCTCTCCGACAGCGCTGAAGGTTGTGTG	1740
Db	2282	ACCTGTCTTGCAACAACGACGATATTTGTGTG	2314

LOCUS	2079 bp	mRNA	linear	PRI 29-SEP-2000
AK026889				
DEFINITION	Homo sapiens cDNA: FLJ23236 fis, clone COL00725.			
ACCESSION	AK026889			
VERSION	AK026889.1	GI:10439854		
KEYWORDS	oligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (sites)			
TITLE	Kawabata, A., Hkiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.			
JOURNAL	NEDO human cDNA sequencing project			
AUTHORS	2 (bases 1 to 2079)			
TITLE	Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.			
JOURNAL	Direct Submission			
COMMENT	Submitted (29-AGO-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)			
FEATURES	<p>1.2079</p> <p>Location/Qualifiers</p> <p>organism="Homo sapiens"</p> <p>molecule="mRNA"</p> <p>db_xref="taxon:9606"</p> <p>clone="COL00725"</p> <p>clone_type="colony"</p> <p>clone_id="COL"</p> <p>note="cloning vector PME18SFL3"</p> <p>304..681</p> <p>note="unnamed protein product"</p> <p>codon_start=1</p> <p>protein_id="BA815584.1"</p> <p>db_xref="GI:10439855"</p> <p>translation="MDDLTFLVDEQFRKGSQSRKRKRGHTASSESDQOWPEBKRLKEDI"</p>			

```

/note="cloning vector pME18SPL3"
304..681
/note="unnamed protein product"
/codon_start=1
/protein_id="BA01584.1"
/db_xref="GI:10439835"
/translation="MODTKRLVGEQKPKGQSRKRGHTVASESDQEQMPEEKRLKEDI

```



```

LENEDEONSPKRRKGRPPKPLGGTPKEEPTWKTSKKSKKKSGPPADEEEER
OSGNTBEKSKSKOHRSVRAOOR"
BASE COUNT      759 a      354 c      339 g      627 t
ORIGIN
Query Match      10.7%; Score 447; DB 9; Length 2079;
Best Local Similarity 85.5%; Pred. No. 1.3e-81;
Matches 550; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

OY 3624 GTCGAATGGAGAGGCTAGAGCAGAAAAAACCCTGCTACAGACAGAGAGGAA 3683
    |||||||
DB 237 GTCGAATGGAGAGGCTAGAGCAGAAAAAACCCTGCTACAGACAGAGAGAA 296
    |||||||
OY 3684 ATTAGTATGATGACTTGAATAAGTTGGTACAGAAACAGAACTTAAGGCACTCAGCG 3743
    |||||||
DB 297 ATTAGTATGATGACTTGAATAAGTTGGTACAGAAACAGAACTTAAGGCACTCAGCG 356
    |||||||
OY 3744 AAGTCGAAAAAGAGCCATACGCGCTTCAGATGACAGACAGTGGCCTGAGAAAA 3803
    |||||||
DB 357 AAGTCGAAAAAGAGCCATACGCGCTTCAGATGACAGACAGTGGCCTGAGAAAA 416
    |||||||
OY 3804 GAGGCTCAAGAGATATATTAGAAATGAAGATGAACAGATAGTCCGCCAAAAAGGG 3863
    |||||||
DB 417 GAGGCTCAAGAGATATATTAGAAATGAAGATGAACAGATAGTCCGCCAAAAAGGG 476
    |||||||
OY 3864 TAAAGAGGCCGACCCACCAAAACCTTTGGTGAAGTACACCAAAAGAGCCACAT 3923
    |||||||
DB 477 TAAAGAGGCCGACCCACCAAAACCTTTGGTGAAGTACACCAAAAGAGCCACAT 536
    |||||||
OY 3924 GAAACTCTTAAAAAGAGCAAAAAAATCTGACCTCCAGCACCAGAGAGAGGA 3983
    |||||||
DB 537 GAAACTCTTAAAAAGAGCAAAAAAATCTGACCTCCAGCACCAGAGAGAGGA 596
    |||||||
OY 3984 AGAAGAGAAAAAGACAAAGTGAATACGAACAGACAGATCCAAAGCAACGACAGT 4043
    |||||||
DB 597 AGAAGAGAAAAAGACAAAGTGAATACGAACAGACAGATCCAAAGCAACGACAGT 656
    |||||||
OY 4044 GTCAGAGAGACACACGACG----- 4062
    |||||||
DB 657 GTCAGAGAGAGACACACGAGTGAAGCATGTGTAACCTTAACGATCTTTTCGTAC 716
    |||||||
OY 4063 ----- 4063
    |||||||
DB 717 TATATTAATAATCAATAATTGATGCTATCCACATTTGGGCTTCCCAAGCAGACGA 776
    |||||||
OY 4071 ATCTCTGAATCTAGTCAATTTGAATCCACACAGTCCACACAGAAAGAGCAGAG 4130
    |||||||
DB 777 ATCTCTGAATCTAGTCAATTTGAATCCACACAGTCCACACAGAAAGAGCAGAG 836
    |||||||
OY 4131 ACCATCAAAAAGGCCATCACCATCAACCAAAAAAATGTG 4173
    |||||||
DB 837 ACCATCAAAAAGGCCATCACCATCAACCAAAAAAATGTG 879
    |||||||

RESULT 14
HS49J10      137246 bp      DNA      linear      PRI 12-OCT-2001
LOCUS
DEFINITION   Human DNA sequence from clone RPI-49J10 on chromosome 13. Contains
              the 3' end of the gene for androgen-induced prostate proliferative
              shut-off associated protein (KIA00979), ESTs, STSs and GSSs,
              complete sequence.
ACCESSION    Z84572
VERSION      Z84572.1 GI:1813974
KEYWORDS     HTG; KIA00979.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 137246)
AUTHORS      Whiteley, M.
TITLE        Direct Submission
JOURNAL      Submitted (11-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

```

COMMENT

requests: clonerquest@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where difference are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>

RPI-49J10 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: PCYFAC2

IMPORTANT: This sequence is not the entire insert of clone RPI-49J10. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RPI-267P19 is at 99 in this sequence.

FEATURES

source

```

repeat_region
    /note="12 repeat: matches 2575. .2706 of consensus"
    /note="1831
    /note="Alusx repeat: matches 1. .281 of consensus"
    /note="1838. .1873
    /note="18 copies 2 mer tg 100% conserved"
    /note="1876. .2848
    /note="TLP15 repeat: matches 5151. .6157 of consensus"
    /note="3149. 3405
    /note="MER4A repeat: matches 1. .247 of consensus"
    /note="3406. .3530
    /note="L2 repeat: matches 2556. .2704 of consensus"
    /note="3815. .3852
    /note="7SLRNA repeat: matches 279. .316 of consensus"
    /note="3853. 3922
    /note="AlusC repeat: matches 120. .188 of consensus"
    /note="4320. .26902
    /gene="49J10.1"
    /product="49J10.1.1 (KIA00979, isoform 1)"
    /note="match: cDNAs: Em:AB023196
    match: ESTs: Em:BF043498 Em:BG571128 Em:BF094231 Em:F12354
    Em:HSC38H11 Em:T64867 Em:BF104916 Em:BB891040 Em:H44781"
    /evidence="not_experimental"
    join(<4320. .4524,6830. .6944,9075. .9207,9522. .9641,
    10616. .10678,11563. .11708,15477. .15582,21108. .21547,
    21638. .21748,24176. .24311,26004. .26902)
    /gene="49J10.1"
    /product="49J10.1.2 (androgen-induced prostate
    proliferative shut-off associated protein, isoform 2)"
    /note="match: cDNAs: Em:U95825"

```

mRNA

CDS
/evidence=not_experimental
join(<4320..4524,6830..6944,9075..9207,9522..9641,
10616..10678,11563..11708,15477..15582,21108..21547,
21638..21748,24176..24311,26004..26039)
/gene="49J10.1"
/note="continued from 267P19.1 in Em:275889
match: proteins: Tr:Q9Y2I5"
/codon_start=1
/evidence=not_experimental
/product="49J10.1.1 (KIAA0979, isoform 1)"
/protein_id="CAC94788.1"
/db_xref="GI:16151354"
/translation="DECYOVQVFAQKLRHGLSRLPLEYMAICALCAKDPYKERRA
HAROCLVKNINVRREYLKQAAVSEKLSLPEYVVPYTHLAHPDVKYVDIOL
KVCECLMFLVLEITAMKNENSHAFIRKVENIKQKQDAGPDDAKMNEKLYVCYA
NMINKSTTYSLESKPDVLPARFTQPKNSNTKNTLPKMSFTYPGKRTTNV
LCAVNPPLSSAGKQSQTKSSRMETVSNASSNSPGRKGLDSDSEMDHSENYT
MSSPLRGKSKDRDSDLVRSLELPRGRKKTPTVEQEKLGMDLTKLVQEQKPKGS
ORSRKRGHASEDEQOMPEERKLKEDILENEDQNSPPKCKGRGPRPLGGGTRE
EPTMKTSKSKSKSGPPAPEEEEROSGNTGROKSKQKHVRRAQOSRAESPES
SAIESQSTPQKGRGPRKTPSPQPKNVNRGRSKQATKENDSSEEDVFGSSPV
DLPDETEEEESTVNVRRRSKREER"
join(<4320..4524,6830..6944,9075..9207,9522..9641,
10616..10678,11563..11708,15477..15582,21108..21547,
21641..21752)
/gene="49J10.1"
/note="continued from 267P19.1 in Em:275889
match: proteins: Tr:Q9Y451"
/codon_start=1
/evidence=not_experimental
/product="49J10.1.2 (androgen-induced prostate
proliferative shut-off associated protein, isoform 2)"
/protein_id="CAC94787.1"
/db_xref="GI:16151353"
/translation="DECYOVQVFAQKLRHGLSRLPLEYMAICALCAKDPYKERRA
HAROCLVKNINVRREYLKQAAVSEKLSLPEYVVPYTHLAHPDVKYVDIOL
KVCECLMFLVLEITAMKNENSHAFIRKVENIKQKQDAGPDDAKMNEKLYVCYA
NMINKSTTYSLESKPDVLPARFTQPKNSNTKNTLPKMSFTYPGKRTTNV
LCAVNPPLSSAGKQSQTKSSRMETVSNASSNSPGRKGLDSDSEMDHSENYT
MSSPLRGKSKDRDSDLVRSLELPRGRKKTPTVEQEKLGMDLTKLVQEQKPKGS
ORSRKRGHASEDEQOMPEERKLKEDILENEDQNSPPKCKGRGPRPLGGGTRE
EPTMKTSKSKSKSGPPAPEEEEROSGNTGROKSKQKHVRRAQOSRAESPES
SAIESQSTPQKGRGPRKTPSPQPKNVNRGRSKQATKENDSSEEDVFGSSPV
DLPDETEEEESTVNVRRRSKREER"
4959..5316
/note="L2 repeat: matches 1830..2229 of consensus"
5608..5810
/note="MIR repeat: matches 46..262 of consensus"
5874..5953
/note="L2 repeat: matches 2625..2736 of consensus"
6757..6826
/note="35 copies 2 mer tt 80% conserved"
7829..8133
/note="AluX repeat: matches 1..297 of consensus"
8892..9087
/note="98 copies 2 mer tt 59% conserved"
10957..11242
/note="AluY repeat: matches 1..291 of consensus"
12344..13119
/note="L1P16 repeat: matches 5321..6156 of consensus"
14108..14427
/note="L2 repeat: matches 2014..2333 of consensus"
14428..14540
/note="HY1 repeat: matches 1..111 of consensus"
14541..14935
/note="L2 repeat: matches 1619..2014 of consensus"
14951..15154
/note="L2 repeat: matches 2529..2730 of consensus"
16579..16885
/note="AluX repeat: matches 4..312 of consensus"
18059..18397
/note="L1P repeat: matches 3809..4138 of consensus"
19076..19481
/note="MLT1A2 repeat: matches 1..374 of consensus"

repeat_region	19502..19807	/note="AluX repeat: matches 1..310 of consensus"
repeat_region	19807..20122	/note="AluY repeat: matches 1..294 of consensus"
repeat_region	23282..23371	/note="L2 repeat: matches 2590..2733 of consensus"
repeat_region	23734..23944	/note="L2 repeat: matches 1814..2050 of consensus"
repeat_region	24795..25085	/note="AluX repeat: matches 1..294 of consensus"
repeat_region	26083..26106	/note="L2 copies 2 mer tt 95% conserved"
misc_feature	26140..26289	/gene="49J10.1"
misc_feature	26749..27206	/note="match: STS: Em:G43328 Em:G43329"
misc_feature	26790..27245	/note="match: GSS: Em:A2019111"
polya_signal	26877..26882	/note="match: GSS: Em:A098541"
polya_site	26899	/gene="49J10.1"
polya_site	26900	/gene="49J10.1"
polya_site	26901	/gene="49J10.1"
polya_site	26902	/gene="49J10.1"
repeat_region	28327..28360	/note="17 copies 2 mer tt 88% conserved"
misc_feature	29194..29299	/note="match: GSS: Em:A0675002"
misc_feature	29212..29650	/note="match: GSS: Em:A0675002"
repeat_region	29300..29554	/note="L2 repeat: matches 2498..2750 of consensus"
Query Match	10.7% Score 447; DB 9; Length 137246;	
Best Local Similarity	85.5% Pred. No. 6e-82;	
Matches 550; Conservative	0; Mismatches 0; Indels 93; Gaps 1;	
QY	3624	GTCTGAATTGGAGAACCTTAGAGGACGAGAAAAACGCCGTCACAGAACAGAGAGAGA 3683
DB	21107	GTCGTGAATTGGAGAACCTTAGAGGACGAGAAAAACGCCGTCACAGAACAGAGAGAGA 21166
QY	3684	ATTAGGTATGATGACTTGACTTAAGTTGGTACAGAAACCTTAAGGCAGTCAGCG 3743
DB	21167	ATTAGGTATGATGACTTGACTTAAGTTGGTACAGAAACCTTAAGGCAGTCAGCG 21226
QY	3744	AAGTCGAAAAGAGGCGCATACGCTCAGAAATCGATGAAACAGCAGTCGAGGAAAA 3803
DB	21227	AAGTCGAAAAGAGGCGCATACGCTCAGAAATCGATGAAACAGCAGTCGAGGAAAA 21286
QY	3804	GAGCTCAAAAGAGATATTAAATAATGAATGAACAGATAGTCGCCCAAAAAAGG 3863
DB	21287	GAGCTCAAAAGAGATATTAAATAATGAATGAACAGATAGTCGCCCAAAAAAGG 21346
QY	3864	TAAAGAGGCGCAGCAACCAACCTTGTGTGAGGTACCAAAAGAGAGCAACAAT 3923
DB	21347	TAAAGAGGCGCAGCAACCAACCTTGTGTGAGGTACCAAAAGAGAGCAACAAT 21406
QY	3924	GAAACTTCTAAAAAAGGAAGCAAAAAAATGTGACCTCCAGCAACAGAGAGAGGA 3983
DB	21407	GAAACTTCTAAAAAAGGAAGCAAAAAAATGTGACCTCCAGCAACAGAGAGAGGA 21466
QY	3984	AGAAAGAAAGCAAAAGTGAATATCGAAATGAAAGATCCAAAGCAACAGCAGAGT 4043
DB	21467	AGAAAGAAAGCAAAAGTGAATATCGAAATGAAAGATCCAAAGCAACAGCAGAGT 21526
QY	4044	GTCAGAGAGACAGCAGCAGC----- 4062
DB	21527	GTCAGAGAGAGACAGCAGCAGCAGTAACTGTAACCTTAACCTGATCTGTTCTTAC 21586

QY 4063 -----AGACGAGA 4070
 Db 21587 TATATTATAATCATATTTGATGCTATCCACATTTGGGTCTTCCCAAGACAGAGAGA 21646
 QY 4071 ATCTCTGAATCTATGCAATTTGATCCACAGACAGCCACACAGAAAGCGAGAG 4130
 Db 21647 ATCTCTGAATCTATGCAATTTGATCCACAGACAGCCACACAGAAAGCGAGAG 21706
 QY 4131 ACCATCAAAAAGCGCATCATCATCAACCAAAAAAATGTG 4173
 Db 21707 ACCATCAAAAAGCGCATCATCATCAACCAAAAAAATGTG 21749

RESULT 15
 AL138820
 LOCUS Human DNA sequence from clone Rp11-380B4 on chromosome 13, complete
 DEFINITION sequence.
 ACCESSION AL138820
 VERSION AL138820.11 GI:11022364
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 168487)
 AUTHORS Smith, M.
 TITLE Direct Submission
 JOURNAL Submitted (01-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT On Oct 25, 2000 this sequence version replaced gi:10715762.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlap as described above.
 This sequence has been finished according to sequence map criteria
 only a small overlap as described above.
 Such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 Rp11-380B4 is from the library RpC1-11.2 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://dcpac.med.buffalo.edu/
 VECTOR: pBACe3.6
 This sequence is the entire insert of clone Rp11-380B4.

FEATURES
 source
 1..168487
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="Rp11-380B4"
 /clone_lib="RpC1-11.2"
 /vector="pBACe3.6"

BASE COUNT 48433 a 33155 c 34624 g 52275 t
 ORIGIN

Query Match 10.7% Score 447; DB 9; Length 168487;
 Best Local Similarity 85.5%; Pred. No. 5.7e-82;
 Matches 550; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 3624 GTCTGAATTTGGAGAAAGCCCTAGAGCGAGAAAAAGCCCGTCAAGAGAGAGAGAA 3683
 Db 20593 GTCTGAATTTGGAGAAAGCCCTAGAGCGAGAAAAAGCCCGTCAAGAGAGAGAA 20652
 QY 3684 ATTAGGTATGATGACTTGTGACTAGTGTGACAGAGAGAGAACTTAAGGAGTACGG 3743
 Db 20653 ATTAGGTATGATGACTTGTGACTAGTGTGACAGAGAGAGAACTTAAGGAGTACGG 20712
 QY 3744 AAGTCGGAAGAGAGCGCATTCAGGCTTCAAGTCTGATGAACAGAGAGAGAGAGAA 3803
 Db 20713 AAGTCGGAAGAGAGCGCATTCAGGCTTCAAGTCTGATGAACAGAGAGAGAGAGAA 20772
 QY 3804 GAGGCTCAAGAGAGATATATTAGAAAAATGAAGATGAACAGAGAGAGAGAGAGAG 3863
 Db 20773 GAGGCTCAAGAGAGATATATTAGAAAAATGAAGATGAACAGAGAGAGAGAGAGAG 20832
 QY 3864 TAAAGAGAGCGACACCAAAACCTCTTGTGTGAGAGTACACCAAAAGAGAGAGAGAG 3923
 Db 20833 TAAAGAGAGCGACACCAAAACCTCTTGTGTGAGAGTACACCAAAAGAGAGAGAG 20892
 QY 3924 GAAACTCTTAAAAAGAGAGCAAAAAAATCTGACCTCCAGCAGAGAGAGAGAGAG 3983
 Db 20893 GAAACTCTTAAAAAGAGAGCAAAAAAATCTGACCTCCAGCAGAGAGAGAGAGAGAG 20952
 QY 3984 AGAAGAGAGAGAGCAAAAGTGAATATCGGAACAGAGAGTCAAAAGCAAGAGAGAGAG 4043
 Db 20953 AGAAGAGAGAGAGCAAAAGTGAATATCGGAACAGAGAGTCAAAAGCAAGAGAGAGAG 21012
 QY 4044 GTCAAGAGAGACACAGCAG-----AGACGAGA 4070
 Db 21013 GTCAAGAGAGACACAGCAGAGAGTGTGTAACCTTAACCTGATCTGTTGCTTAC 21072
 QY 4063 -----AGACGAGA 4070
 Db 21073 TATATTATAATCATATTTGATGCTATCCACATTTGGGTCTTCCCAAGACAGAGAGA 21132
 QY 4071 ATCTCTGAATCTATGCAATTTGATCCACAGACAGCCACACAGAAAGCGAGAGAG 4130
 Db 21133 ATCTCTGAATCTATGCAATTTGATCCACAGACAGCCACACAGAAAGCGAGAGAG 21192
 QY 4131 ACCATCAAAAAGCGCATCATCATCAACCAAAAAAATGTG 4173
 Db 21193 ACCATCAAAAAGCGCATCATCATCAACCAAAAAAATGTG 21235

Search completed: September 24, 2003, 14:13:31
 Job time : 14881 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:22:05 ; Search time 11608 Seconds

(without alignments)
4902.248 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRNDCKITYPGVC.....OKGRGRPSKTPSPQPKNV 1391

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US09512581/runat_23092003_163546_5474/app_query.fasta_1.1543
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=humand0.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=eto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09512581.ecgn.1.1.7880_etunat_23092003_163546_5474 -NCPV=6 -ICPV=3
-NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -XGAPOP=6
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_in:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7155	99.5	5309	9	AB023196	AB023196 Homo sapi
2	7155	99.5	7444	9	HS26H231	AL137201 Novel hum
3	7101	98.7	5253	9	HS095825	U95825 Human andro
4	6942	96.5	5317	10	AY102267	AY102267 Mus muscu
5	5659	78.7	5944	10	AK122414	AK122414 Mus muscu
6	4760.5	66.2	6744	9	AF294791	AF294791 Homo sapi
7	2953.5	41.1	1852	9	HS050533	U50533 Human BRCA2
8	2704.5	37.6	5177	9	AB014548	AB014548 Homo sapi
9	2517	35.0	4252	9	BC039256	BC039256 Homo sapi
10	2474	34.4	2265	9	AK098331	AK098331 Homo sapi
11	2474	34.4	2705	9	BC041361	BC041361 Homo sapi
12	2442	33.9	1976	10	BC021408	BC021408 Mus muscu
13	1836	25.5	216608	5	AL844559	AL844559 Zebrafish
14	1729.5	24.0	173613	3	AC007475	AC007475 Drosophill
15	1729.5	24.0	192763	3	AC007474	AC007474 Drosophill
16	1729.5	24.0	292919	3	AB003823	AB003823 Drosophill
17	1708.5	23.8	194634	2	AC020286	AC020286 Drosophill
18	925.5	12.9	2079	9	AK026889	AK026889 Homo sapi
19	925.5	12.9	137246	9	HS49J10	Z84572 Human DNA s
20	925.5	12.9	168487	9	AL138820	AL138820 Human DNA
21	840.5	11.7	3779	9	BC009650	BC009650 Homo sapi
22	833.5	11.6	4046	8	AB067651	AB067651 Schizosac
23	833.5	11.6	9982	8	SPAC110	AL441624 S.pombe C
24	819.5	11.4	54398	2	AC068224	AC068224 Homo sapi
25	769	10.7	236508	2	AC111126	AC111126 Mus muscu
26	768	10.7	72157	2	AC016449	AC016449 Homo sapi
27	747	10.4	263479	2	AC125986	AC125986 Rattus no
28	706.5	9.8	2212	6	BD159124	BD159124 Primer fo
29	706.5	9.8	2212	9	AK021757	AK021757 Homo sapi
30	698	9.7	439	6	BD059747	BD059747 Secreled
31	697.5	9.7	6411	8	SMA99344	AJ009934 Sordaria
32	645	9.0	772	6	BD145886	BD145886 Primer fo
33	635	8.8	4885	8	EMEBIMD	L03300 Emericella
34	584	8.1	153751	3	AC116551	AC116551 Dictyoste
35	554.5	7.7	54398	2	AC068224	AC068224 Homo sapi
36	523	7.3	184321	2	AC110493	AC110493 Mus muscu
37	523	7.3	190058	2	AC112263	AC112263 Mus muscu
38	523	7.3	298053	2	AC116574	AC116574 Mus muscu
39	520	7.2	233589	2	AC095342	AC095342 Rattus no
40	512	7.1	163355	2	AC117927	AC117927 Mus muscu
41	505	7.0	1541	3	AY058338	AY058338 Drosophill
42	501	7.0	295	6	BD026562	BD026562 Sequence
43	466	6.5	98856	9	AC113150	AC113150 Homo sapi
44	466	6.5	145450	2	AC068352	AC068352 Homo sapi
45	401.5	5.6	85592	2	AL512630	AL512630 Mus muscu

RESULT 1

ALIGNMENTS

AB023196 5309 bp mRNA linear PRI 10-MAY-2002
LOCUS Homo sapiens mRNA for KIAA0979 protein, partial cds.
DEFINITION AB023196
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
1. 5309
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hj07056s1"
/sex="male"
/tissue_type="brain"
/clone_lib="pluescriptII SK plus"
/dev_stage="adult"
/note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 722 was derived from fj09522 and 723 - 5309 was derived from hj07056)."
1. 5309
/gene="KIAA0979"
/gene="KIAA0979"
/codon_start=1
/protein_id="AA7683.2"
/translation="GGGRRRLDAAAADPGEPRSESSSATSRRGNI SVANSKTRT NDGKITPPGVKEISDKISKEEMVRRLKMYVKTMDMDQDSEEEKLYNLALIASD FFLKHPDKVALLAACLDIFRIYAEAPTSDDKLDKIDMFITROLGLEDKSDQ ENRFYLLLENIAWVKSNYCELDNSMIEFOLYRTLEFSVNNHNGKHVHMADLMS SITCEGTVSOGELIDYLVNLYPAHKNLNOAYLAKALKRLTKQAIEPYITTFNFOY LMAKRTISDSSEHVPDILIELINIDSHLSVLPLEFKLSNDNERLDVAVLAK MFGAKDSELASQNPWQCYIGRPNDIHVPRLCEVFAISCLNHDPLADILEYLA VRSHPDEAARHVDYIVISYPAKKDILLNDHNLVFEERTLDKRWYRKANMGLAO IYKKYALOSAAKDAKOIAIMIKDLHIYONSIDRLIVERFAQYVMYAINETTE RMKCLLYLATLIDNNAVKALNEMKCONLIRHOVKLIDLIDIKOPKTASVAINETSKM VITRNLPDPSKADPMKPTVOVLEDDKIRKOLREVLYSPSCSCQAEGCVRETTKAG MPKOPTNPFLEMLIFELERLAPVHIDTESIALIKOVNKSIDGTADDEDEVPTDQAL RAGLELLKVLSTFHPISFSAETESILACIMDEKVEVAEALQIFNTGSKIIEDEP HIRSALELPVLHHSKSKGPPOAKYAIIHICIAIFSSKEFOEQIPEPLHKSIDPNSLEH LITPLVIGHIALLAPOFAAPLKSIVATFIVKDLMDNDRLPGKRTKLWYDDESVS ETWYKQIATIMMWVWILGKMNHSGSTIRLLTTLISGDLTBOGKTSKPMBSRL RLAAGSAIVKLAOPRCYHEITITTEYOYLCALINDECYQVNOYVAQGLHKLSRLRLP LETHAICALCAKDPKERRARAOCLKINIVREYIKNOHSAVEKLSLLEPVEVAPY THILHADPDYVKVODLEOLDKVEKCLMKNIVREYIKNOHSAVEKLSLLEPVEVAPY AOGDDAKMNEKLTVCYDAMNIIIMSKTSTYLSGPKDPVLPARFOPDPNENSTKN YLPEEMKSFTEPKPTKTYNAGVNRPLSSGKOSOTKSSMETEVSNASSNSPSG RIGRLDSSEHSDENEDYTMSSPLPGKRSKRDSDLVRSLEKLPGRKKTPTVBOE EKIAMDLTITLVQOKRPGKSORSKRKRHTKTSASESDQOPEPKRLKEDILEDNDSNP PKGKRGRRPPPLGGTPEKPEPTKTKSGSKKSGPAPAEPEEEERESGNTGSKK

SKORVSRRAQORAESESSAIESSTOSTPOJGRGRPSKTPSPQKKNVRGRSKOAA
TKENDSSSEVDVFOGSSPVDDIPQEEFEEREEESTVNNRRSRKRRR"
BASE COUNT 1774 a 955 c 1115 g 1465 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 5309
Score: 7155.00 Matches: 1387
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 4
Query Match: 99.47% Indels: 0
DB: 9 Gaps: 0
US-09-512-581b-2 (1-1391) x AB023196 (1-5309)
QY 1 MetLAHISserLySThArgThrASnAspLyLysIleThrYrProProGlyValLys 20
DB 109 ATGCCTCATTTCAAGACTAGAGCAATCATGTAATAATACATATCCCGGGGCTCAAG 168
QY 21 GluIleSerAspLysIleSerLySGluIleMetValArgArgLysMetValValLys 40
DB 169 GAATATCATGATTAATAATATCTAAAGAGAGAGTGGTGAACGATTAAAGATGTTGTAAA 228
QY 41 ThrPheMetAspMetAspGlnAspSerGluIleLysGluIleuTyrrLeuAsnLeuAla 60
DB 229 ACTTTTATGATATGACACGACGACTGCAAGAAAGAGAGCTTTATTTAAACCTAGCT 288
QY 61 LeuHISLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
DB 289 TTACATCTTGCTTGAGATTTTCTTCAAGCATCTCTGAAAGATGTTCTTACTGCTGTA 348
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrrAlaProGluAlaProTyrrThrSerPro 100
DB 349 GCCTGCTGCTTGGTGTGATATTTTCAGAGATTATCTCTGAGGCTCCTTACACATCCCT 408
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGluLeuLysGlyLeuGluAspThr 120
DB 409 GATTAACATTAAGATATATTTATTTATTAACAAGACAGCTGAAGGGCTGAGAGATACA 468
QY 121 LysSerProGlnIlePheAsnArgTyrrPheTyrrLeuGluGluAsnIleAlaIrrValLysSer 140
DB 469 AAGAGCCCAATTCATATAGTATTTTATTTACTTGAGACATCTGCTGGGTCAAGTCA 528
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrrArgThr 160
DB 529 TATAACATATGCTTGTGATTAAGATATGACATGAATAATTTCCACCAGCTTATACAGAAC 588
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
DB 589 TTATTTTCAGTTAATAACAAATGGCCACAAATCAGAAAGCCATATGACATGCTGAGACCTT 648
QY 181 MetSerSerIleIleCysGluGluAspThrValSerGlnGluLeuAspThrValLeu 200
DB 649 ATGACCTCTATATTTTGTGAAGGTGATACACTGTCTCAGGAGCTTTTGGATACGTTTGA 708
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrrAspLeuAlaLysAla 220
DB 709 GTAATCTGTAACCTGCTCATTAAGATTAATAACAAGACATATGATTTGGCAAGGCT 768
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrrIleThrPhePheAsnGlnVal 240
DB 769 TTACTGAAGAGAGACGCTCAAGCATATGACGCAATATATACCAATTTTAAAAACAGGTT 828
QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValIlePheAspLeuIleLeu 260
DB 829 CTGATGCTGGGAAACAATCTATACGCGATTTTGACAGGCACTGCTTTGACTTAATTTTG 888
QY 261 GluLeuTyrrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
DB 889 GAGCTTCACATATTAATGATGATGCTGCTGCTGTTTAAACCCAGCTTGAATTTAAA 948
QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300

Db 949 TTAAGAGCAATGATTAATGAGGAGCCCTACAGTTGTTAACTACTGCGCAAAATGTTT 1008
QY 301 G1A1A1yAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnGlySerLeuGly 320
Db 1009 GGGGCAAAAGGATTGCAATGGCTTCTCAAAACAAAGCCATTGGCAGAGCTACCTGGGC 1068
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
Db 1069 AGGTTAATGATATCATCATCTACCATCCCGCTGGAAATGTGTCAAAATTTCTGACCATTT 1128
QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
Db 1129 CTCATGAACCATCCGATTACCAAAAGCCTTAACAGACTATCTTAAATGAGGCACAT 1188
QY 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
Db 1189 GACCTGAGGAAGCTATTAGACATGATGTTATTGTCTCATATGTTACACCTGCTCAAAAG 1248
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
Db 1249 GATATTCTTCTGCTCAATGATCATCTTACTTAATTTTGTGAGAGAGAACATTAAGCANA 1308
QY 401 ArgTrpArgValArgLysGluAlaMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
Db 1309 CGATGGAGAGTACGCAAAAGAACCATGATGAGCCTTGCCCAATTTATAGAAATATGCT 1368
QY 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
Db 1369 TTACAGTCAAGCTGGAAAGATGCTGCAAAACAGATGATGATGCTCAAAAGCAATTTG 1428
QY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
Db 1429 CTACATATATATATCAAAATAGTATGATGATGATGATGATGATGATGATGATGATGAT 1488
QY 461 GlnTyrMetValProHisAsnLeuGluThrGluArgMetLysCysLeuTyrTyrLeu 480
Db 1489 CAATCATGATGCTCCACAAATTTAGAAACTACAGAACCGATGAAATGCTTATATATCTTG 1548
QY 481 TyrAlaTrpLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
Db 1549 TATGCGACAGTGGATTTAAATGCTGTGAAGCATGGAATGGAATGGGAAATGCTCAAAAT 1608
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
Db 1609 CTGCTCCGACATCAAGTAAGAATTTGCTGACTTGATTAACCAACCCAAACAGATGCC 1668
QY 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
Db 1669 AGTGTCAAGGCCATATTTTCAAAAGTGAATGATGATTAACAAGAAATTTACCTGATCCTGCT 1728
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
Db 1729 AAGGCTCAGAGATTATGAAGAAATTCACACAGGTGTAGAGATGATAGAAATTAACA 1788
QY 561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluLysCysVal 580
Db 1789 AAGCATTAAGAGTACTGTGTGTCACACATGCTCTCTGCAACAGAGCTCAAGGTGTGTG 1848
QY 581 ArgGluIleThrLysLysLeuGluLysProLysGlnProThrAsnProPheLeuGluMet 600
Db 1849 CGTGAATTAACATAAGTTGGGCAACCCCAAAACAGCTACAAATCTTCTGGAAGATG 1908
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
Db 1909 ATCAAGTTTCTCTTGAGAGATACACCTGTGCACATAGATACCGAAATCTATCAGTGT 1968
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluLysVal 640
Db 1969 CTTATTAACAAGTGAACAATCAATAGATGAGACAGCAGATGAGATGAGATGAGATGAGT 2028
QY 641 ProThrAspGlnAlaIleArgAlaGluLeuGluLeuLysValLeuSerPheThrHis 660
Db 2029 CCAACTGATCAAGCCATCAGACAGGTCTTGAAGTGTAAAGTACTCTCATTTTACACAT 2088

QY 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
Db 2089 CCCATCTCATTTTCATTTCTGCTCAAAACATTTGAATCTTACTGCTTGTGTGAAATGAT 2148
QY 681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2149 GATGAAGAAAGTACAGAAACCTGCACATAATTTTCAAAACACAGAGAAAGCAAAATTTGA 2208
QY 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLys 720
Db 2209 GAGGATTTTCCACACATCAGATCAGCTTGTCTTCTGTTTACATCACAAATCTAAGAAA 2268
QY 721 GlyProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLys 740
Db 2269 GAGACCCCGCTCAACGCCAAATATGCTCATTTCTGATCATGCGCATATTTCTAGTAAA 2328
QY 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2329 GAGACCCAGTTTGCACAGATATTTGAGCCTGTGCATTAAGAGCTAGATCCAAAGCAACTG 2388
QY 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2389 GACATCTCATTAACACCATTTGCTACTATTTGCTCATATTTGCTCTGCTGACCTGATCA 2448
QY 781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2449 TTTGCTGCTCTTGAATCTTTGGATGCTTACTTTTCTATTTGTAAGATTTTCTCATGAT 2508
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluLysValSerPro 820
Db 2509 GATCGGCTTCCAGGGGAAAAAGAACACTAAACTTTGGGTTCCAGATGAAGACATCTCT 2568
QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGluMetLys 840
Db 2569 GAGACAAATGCTCAAAATTCAGCTATTAAGATGCTTGCATGCTGCTGCTGCTGCTGCT 2628
QY 841 AsnAsnHisSerLysSerGlyThrSerTrpLeuArgLeuLeuThrIleLeuHisSer 860
Db 2629 AATTAATCAGATTAATCAGGAACCTTCTACCTTAAGATTTCTTAACAATATTTGCTATG 2688
QY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2689 GATGGAGACTTACAGAACAGGGAAGAAATTAAGTAACAGATATGACAGCTGAGACTT 2748
QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleThr 900
Db 2749 GCTGCTGGAGTGTATTTGTGAAGCTGGCACAAGAACCTGTTTACCATGAAATCAACACA 2808
QY 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Db 2809 TTAGAACATATATCAGTATGCTATGCTATTAACAGATGAATGCTATTAAGTAAAGACAA 2868
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluLysIleMet 940
Db 2869 GTGTTTGCCCAAGAACTTCAACAAGGCCCTTCCCGTTTACGGCTTCCACTGAGTATAG 2928
QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Db 2929 GCAATCTGTGCTTGTGTCACAAAGATCTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2988
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
Db 2989 TGTGTGGTGAAGAAATTAATATTAAGGCGGAGTATCTGAACAGCATCAGCTGTATAT 3048
QY 981 GluLysLeuLeuSerLeuLeuProGluTyrValAlaProTyrThrIleHisLeuLeuAla 1000
Db 3049 GAAAAATATTTCTCTTACACAGATATGTTGTTCCATATTAACAATTCACCTTTGGCA 3108
QY 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3109 CATGACCCAGATATGCTCAAAAGTATACAGATATTAAGCAACCTTAAGAGTGTAAAGATGT 3168

QY	1021	LeuThrPrpPheValLeuGluIleLeuMetAlaLysSngluAsnAsnSerHisAlaPheIle	1040
Db	3169	CTTGGTTGGTCTCGGAATATTATATGGCTTAAATGAATAATTAACAGTCACGGCTTTATTC	3228
QY	1041	ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAlaLys	1060
Db	3229	AGAAAGATGATGAGAAATATTAAACCAACAAAGATGCCCAAGACCGATGATGCCAAA	3288
QY	1061	MetAsnGluLysLeuTyrrThrValCysAspValAlaMetAsnIleIleMetSerLysSer	1080
Db	3289	ATGATGATGAAAAACGTACACTGCTGTGTGATGTTCCATCATATCATCATGCAAGAGT	3348
QY	1081	ThrThrTyrrSerLeuGlnLysProLysAspProValLeuProAlaArgPhePheThrGln	1100
Db	3349	ACTCATATCAATTTGGAAATCTCTTAAACCCGGTACTACCAAGCTGTTTCTTCACTCA	3408
QY	1101	ProAspLysAsnPheSerAsnThrLysAsnTyrrLeuProGluMetLysSerPhePhe	1120
Db	3409	CCTGACAAAGATTTCAGTAAACCAAAAATTATCTGCTCCTCGAATGCAATCATTTTTC	3468
QY	1121	ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer	1140
Db	3469	ACTCCTGGAAAACTTAAACCAACCAAGTTGTTAGAGCTGTTTAAACAAGCCCTTTCATCA	3528
QY	1141	AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAspAlaSerSer	1160
Db	3529	GCAGGCAACGATCTCAGACCAAAATCATCAGAAATGGAACTGTAGCAATGAACAGC	3588
QY	1161	SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp	1180
Db	3589	AGCTCAATTCGAAGTCTCTGGAAGATTAAGGGAGGCTTGATAGTTCGAAATGGAT	3648
QY	1181	HisSerGluAsnGluAspTyrrThrMetSerSerProLeuProGlyLysLysSerAspLys	1200
Db	3649	CACAGTGAATAATGAAGATTACACATGCTCTCACCTTGGCCGGGGAAAAAAGACAGACAG	3708
QY	1201	ArgAspAspSerAspLeuValaIrisSerGluLeuGluLysProArgGlyArgLysLysThr	1220
Db	3709	AGAGACGACTGTGATCTTGTAAGGCTGAATTGGAGAAAGCCTTAGACGCGAGAAAAACG	3768
QY	1221	ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu	1240
Db	3769	CCCGTCACAGAAACAGAGAGAAATTTAGTGTGATGACTTGACTAAGTTGCTACAGCAA	3828
QY	1241	GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp	1260
Db	3829	CAGAAACCTTAAGGCAAGTCACAGCAAGTCGAAAAAGAGCCATACGGCTTCAGAAATCTGAT	3888
QY	1261	GluGlnGlnTrpProGluLysLysArgLeuLysGluLysPheLeuGluAsnGluAspGlu	1280
Db	3889	GACACGCACTGGCTCGAGAAAAGAGGCTCAAGAAAGATATATTAAGAAAAAGAAAGATGAA	3948
QY	1281	GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly	1300
Db	3949	CAGAAATAGTCGGCCAAAAAAGGTTAAAGAGCGGACCCACCAAAAACCTTGTGGAGCT	4008
QY	1301	ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerLys	1320
Db	4009	ACACCAAAAGAGAGCCCAACATGAAAACTCTTAAAAAAGACCAAAAAAATCTGGA	4068
QY	1321	ProProAlaProGluGluGluGluGluGluGluLysArgGlnSerGlyAsnThrGluGlnLys	1340
Db	4069	CCTCCAGCACGAGAGAGAGAGAAAGAAACAAACAACTGGAAAAATCCGAACAGAG	4128
QY	1341	SerLysSerLysGlnHisArgValSerArgArgIleGlnGlnAsnArgAlaGlnSerProGlu	1360
Db	4129	TCCAAAAGCAACACGACCGAGTGTCAAGGAGACACGACGAGAGCGAATTCCTGAA	4188
QY	1361	SerSerAlaIleGlnSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys	1380
Db	4189	TCTAGTGAATTTGAATCCACACACTCCACACACAGAAAGACGAGAGAGCCATCAAAA	4248
QY	1381	ThrProSerProSerGlnProLysLysAsnVal	1391

Db 4249 AGGCATCATCATTCAACCAAAAAAATGTC 4281
 |||||||
 RESULT 2
 LOCUS HS26H231 7444 bp mRNA Linear PRI 25-JAN-2000
 DEFINITION Novel human gene mapping to chromosome 13.
 ACCESSION AL137201
 VERSION AL137201.1 GI:6759511
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 7444)
 AUTHORS Rhodes, S. and Huckle, E.
 TITLE Direct Submission
 JOURNAL Submitted (13-JAN-2000) E-mail contact: humgeny@sanger.ac.uk
 COMMENT This cDNA sequence was assembled from public domain ESTs and single
 pass sequencing reads from expressed DNA templates, aligned to the
 genomic DNA sequence from the bacterial clones 26H23 (284467),
 26P19 (275869) and 49J10 (284572).
 The EST sequences listed match this sequence with an identity of at
 least 95% between the coordinates shown.
 Further information can be found at
 http://www.sanger.ac.uk/Help/Ch13/ Experimentally determined gene
 Sanger Centre name : 26H23.Cl3.1.
 Location/Qualifiers
 1..7444
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="13q12.3"
 1..107
 /number=1
 108..234
 /number=2
 127..4470
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CA869911.1"
 /db_xref="GI:6759512"
 /translation="MAHSKTRINDGKITYPGKVEISDKSKEBMARLKMYVKTEND
 MDQOSEERKELYLNLAHLASDFELKHPDKDVLVACCLADIFRIYAPAPYSPDK
 LKIDFMEITRLQKGLDSTKSPENRYEYLLENIAVMYSYICCELEPESNIFQLYKT
 LFSEYNNHONKVMHMYNDLMSSIIICGDIYSGDLDTVLVNLVPAKNNLKQAYDIA
 KALKRTAQTAEPIYTNFPNVLMLGKTSISDSEHEDFLLILEYINDSHLLSVLPQ
 LEFKNSNDNERQYVYKILAKMGAGDSELASNNKRLNOCYIGRNDHVPRLREY
 RFASCHLNHPDLAKIDLEVLKVSHPDEERIRDVIVSVTAKKRIILVNDHLNF
 VREFTLDRWRVRKEEAMGELAQIYKRYKLAAGDAKQIAWIKDKLHLYNDLSD
 DRLIVERIFAQYVPHNLLETTERMKCYIYATDILNAVALNEMKCOMLHYOVD
 LLDILDKPDTASVKAIFSKMYVTNRLLPYGAKQDMPKKEFTYQLEDEKIRKOLEV
 VSPKSGQKGAEGVREITTKIKGNKOPTNPLEMIFKLLERIAVPHIDTMSIALIQ
 VNKSIDSTADDEDCGVPDQAIRAGLLILAVLSTHPISSHATPESLLACLAKMDE
 KVAFAAIQIFKNTGSKIIEDPPIHRSLLPEVLHKSKGPPROKRTAICHTIAIFSK
 ETQQAQIIFKPLHSLDPSNLEHLITPLVTIGHIALADPFAAPLKSIVATFVYKDL
 MNDLPGKTKTKLVNPDEEVSPEYMWIOAKIMVRYMLLKMNNHSGSGSTRLLTLIT
 ILHSDGTEDEGKISKDMSRLRLAAGSAIVYKLAQECYENHITLDEYOCLALINDE
 CYQVRYPAOKLHKGLSRLRLPLEYMAICALCKMPKEERARAQCLVANNVAREY
 LKQNAVSEKLSLLPREVVPYTHLADHPDYKQVODIEQLQVCKCLFVLEITMA
 KNEVNSHAFIRKVENYIKOTKDAQGPDAKAMNEKLYVQCVANMINKSTTYSLSP
 KDPLPFAFTFQDPKNSNTKNYLPPMKSEFTPGKPKTNTVLAVNAPLSSGKQSQ
 TKSSEMELEVNASSSSNPSPGRIGKGLDSEMDHSENEYDTNMSPLPKGSPKRODS
 DLYSSELEKPRGRKTPVTEQEEKLMDGLTKLYQROKPGSGORSRGRGHTASRSDQ
 OMPEERKLKEDIENEDONSPPKKCKGRPPKPLGGCTGKREPTTKTSKSKGSKKSG
 PPAPEESEEEROSGNTQKSKOHVNSRAQGRASPSASILESTQSPQKGRGP
 SKTSPSPQKKNVGRSKQATKSEVSDVFGQSSVPDIDIPQETEEVEEYSTVN
 VRRSAKRRR"
 235..438
 /number=3
 439..525
 /number=4

```

exon          526..623
              /number=5
exon          624..750
              /number=6
exon          751..831
              /number=7
exon          832..972
              /number=8
exon          973..1088
              /number=9
exon          1089..1183
              /number=10
exon          1184..1329
              /number=11
exon          1330..1481
              /number=12
exon          1482..1595
              /number=13
exon          1596..1677
              /number=14
exon          1678..1726
              /number=15
exon          1727..1866
              /number=16
misc_feature  1866..2046
              /note="matches EST AA687846 from clone IMAGE:1219676"
misc_feature  1866..1941
              /note="matches EST AA687776 from clone IMAGE:1219628"
exon          1867..1982
              /number=17
exon          1983..2088
              /number=18
exon          2089..2249
              /number=19
exon          2250..2373
              /number=20
exon          2374..2532
              /number=21
exon          2533..2601
              /number=22
exon          2602..2738
              /number=23
exon          2739..2862
              /number=24
exon          2863..3067
              /number=25
misc_feature  3067..2912
              /note="matches EST AA630616 from clone 856164"
exon          3068..3182
              /number=26
exon          3183..3315
              /number=27
misc_feature  3302..3572
              /note="matches EST AA333118"
exon          3316..3435
              /number=28
misc_feature  3349..3818
              /note="matches EST AA282190 from clone IMAGE:704631"
exon          3436..3498
              /number=29
exon          3499..3644
              /number=30
misc_feature  3644..3586
              /note="matches EST A1655429 from clone IMAGE:2239331"
exon          3645..3750
              /number=31
exon          3751..4190
              /number=32
misc_feature  4190..3864
              /note="matches EST A1680124 from clone IMAGE:2264521"
misc_feature  4001..4119
              /note="matches EST N83765 from clone KK2745"
misc_feature  4135..4450

```

```

exon          /note="matches EST F12354 from clone C-38h11."
              4191..4298
              /number=33
misc_feature  4255..4307
              /note="matches EST T64867 from clone 22264"
exon          4299..4434
              /number=34
misc_feature  4398..4533
              /note="matches EST H44781 from clone 188032"
exon          4435..7444
              /number=35
misc_feature  4537..5022
              /note="matches EST A1681422 from clone IMAGE:2272632"
misc_feature  4537..4892
              /note="matches EST AA911125 from clone IMAGE:1416862"
misc_feature  4538..4969
              /note="matches EST A1275866 from clone IMAGE:1877482"
misc_feature  4538..4947
              /note="matches EST AA987361 from clone IMAGE:1603350"
misc_feature  4538..4943
              /note="matches EST A1299646 from clone IMAGE:1898010"
misc_feature  4562..5003
              /note="matches EST AA447433 from clone 784555"
misc_feature  4571..4854
              /note="matches EST H93424 from clone 220700"
misc_feature  4571..4649
              /note="matches EST W20070 from clone 306049"
misc_feature  4746..4994
              /note="matches EST AW089775 from clone IMAGE:2593146"
misc_feature  4746..4994
              /note="matches EST AA747568 from clone IMAGE:1269696"
misc_feature  4803..5332
              /note="matches EST A1911784 from clone IMAGE:2329873"
misc_feature  4828..5296
              /note="matches EST AA563884 from clone IMAGE:1013990"
misc_feature  4861..5243
              /note="matches EST D20082 from clone pml283"
misc_feature  4872..4921
              /note="matches EST AA177347 from clone 621810"
misc_feature  4872..4921
              /note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature  4872..4921
              /note="matches EST AA154932 from clone 599194"
misc_feature  4872..4921
              /note="matches EST AW048929 from clone
              UT-M-BH1-amo-C-09-0-UT"
misc_feature  4876..5333
              /note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature  4922..5281
              /note="matches EST A1012649 from clone RPLBNO2"
misc_feature  4922..4975
              /note="matches EST AA899586 from clone
              UT-R-E0-cy-e-02-0-UI"
misc_feature  4931..5332
              /note="matches EST AA625961 from clone 745478"
misc_feature  4951..5332
              /note="matches EST A1140790 from clone IMAGE:1566222"
misc_feature  5021..5309

```

Alignment Scores:

```

Pred. No.:      0
Score:          7155.00
Percent Similarity: 99.71%
Best Local Similarity: 99.71%
Query Match:    99.47%
DB:             9
Gaps:           0

```

US-09-512-581B-2 (1-1391) x HS26H231 (1-7444)

```

QY      1 MetAlaHhSserLyThrArgThrAsnAspGlyIleHrTyrProProGlyValLys 20
|||||
Db      127 ATGGCTCATTTCAAGACTAGGACCAATGATGAAAAATTACATATCCGCTGGGGTCAAG 186

```


QY	21	GIUIEserApLySIeSerLySGLuIuMeValArGaLeuLyMeValValLyS	40
Db	187	GAATATCAAGTAAATATCTAAAGAGAGAGATGCGAGACATTTAAATGATGTTCTGAAA	246
QY	41	ThrPheMeLAspMeLAspGlnAspSerGluGluGlyLysGluLeuTyLeuAsnLeuLa	60
Db	247	ACTTTATGATATGAGACCGAGACTCTAGAGAGAAAAGAGCTTATTTAAACCTACT	306
QY	61	LeuHISLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArGLeuLeuVal	80
Db	307	TTACATCTGGCTGAGATTTTTTCTCAAGCATCTGTATAAGATGTTGCGTTACTGTA	366
QY	81	AlaCysCysLeuAlaAspLlePheArgLleTyAlaProGluAlaProTyTrhSerPro	100
Db	367	GCGTGCTCCCTGGATGATTTTTCAGATTATGCTCTGAAGCTCCCTTACACATCCCT	426
QY	101	AspLysLeuLysAspLlePheMerPheLleThrArgGlnLeuLysGlyLeuGlnAspThr	120
Db	427	GATAACCTAAAGGATATATTTATGTTTAAACAGACAGTTGAGGGGCTAGAGATACA	486
QY	121	LysSerProGlnPheAsnArgTyPheTyTrLeuLeuGlnAsnLleAlaTrpValLysSer	140
Db	487	AAGAGCCCAATTCATTAAGATTTTATTTACTTGAGAACATGCTTGGGTCAAGCA	546
QY	141	TyrAsnLleCysPheGlnLeuGlnAspSerHisGluIlePheTrhGlnLeuTyArGthr	160
Db	547	TATATCAATATCTTGGATTTGAAAGATVAGCAATGAAATTTTCACCACCTATACGAACC	606
QY	161	LeuPheSerValLleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu	180
Db	607	TTATTTTCAGTTATTAACAATAGCGCACATCAGAAAGTCATATGCACATGCTGACCTT	666
QY	181	MetSerSerLleLleCysGluGlyAspThrValSerGlnGluLeuAspTrhValLeu	200
Db	667	ATGAGCTCTATTTATTTGTGAAGGTCATACAGTCTCTCAGAGAGCTTTTGATACGGTTTA	726
QY	201	ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyArAspLeuAlaLysLa	220
Db	727	CTAAATCTGGTACCTGCTCATATAGCATTTTAAACAAGCAAGATATGATTTGGCAAAAGCT	786
QY	221	LeuLeuLysArgThrAlaGlnAlaLleGluProTyTrlIethTrhPhePheAsnGlnVal	240
Db	787	TTACTGAAGAGACAGCTCAAGCTATTAGCCATATATATACCATATTTTAAATCAGGTT	846
QY	241	LeuMetLeuGlyLysTrhSerLleSerAspLeuSerGluHisValPheAspLeuLleu	260
Db	847	CTGATGCTTGGGAAAAACATCTATCATCAGCATTTGTGCAGAGCATGCTTTGACTTAATTTTG	906
QY	261	GluLeuTyTrAsnLleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlnPheLys	280
Db	907	GAGCTCTACATATTTGATGATGATTTGCTGCTCTGTGTTTAAACCCACCTTGAATTTAAA	966
QY	281	LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMerPhe	300
Db	967	TTTAAAGAGCATGATTAATGAGAGAGCGCTACAAAGTTGTTAACTAGTGGCAAAAATGTTT	1026
QY	301	GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyTrGlnCysTyTrLeuGly	320
Db	1027	GGGGCAAAAGGATTCAGAAATTTGGCTTCTCAAAAACAAGCACATTTGGCAGCTGACTTGGGC	1086
QY	321	ArgPheAsnAspLleHisValProLleArgLeuGlyCysValLysPheAlaSerHisGys	340
Db	1087	AGGTTATATGATATCCATGATACCAATCCGCTGGCAATGTGAAATTTGCTAGCCATTTGT	1146
QY	341	LeuMetAsnHisProAspLeuAlaLysAspLeuTrhGluTyTrLeuLysValArgSerHis	360
Db	1147	CTCATGAACCATCTCGATTTAGCAAAACACTTAACAAGATGTTCTTAAAGTGAAGTCACAT	1206
QY	361	AspProGluGluAlaLleArgHisAspValLleValSerLleValTrhAlaAlaLysLys	380
Db	1207	GACCTGAGAGAGATATAGACATGATGATTTATGTGTCATAAGATTACAGCTGCTAAAAG	1266

QY	361	ASPILEuleuleuValaAsnAspHisleuleuAsnPhelaIargGIuaRThleuAspLys	400
Db	1267	GATATTTCTTCGGTCAATGATCACCATTACTTAATTTTGGAGAAGACAACATTAGCAAA	1326
QY	401	ArgTPArgValaArgLysGIuaIaMetetGlyLeuAlaGlnIleTyrLysLysTyrLa	420
Db	1327	CGATGGAGAGTACGCCAAGAAAGCCATTGATGGACTTGGCCAAATTTATAGAAATATCT	1386
QY	421	LeuGlnSerAlaIaIaGlyLysAspAlaAlaLysGlnIleAlaTyrIleLysAspLysLeu	440
Db	1387	TTACAGTCAGACGCTGGAAGAAGATGCTCCAAACAGATAGCATGATCAAGACAATTTG	1448
QY	441	LeuHisIleTyrTyrGlnAsnSerIleAsnAspArgleuleuValaGIuaRlIephela	460
Db	1447	CTACATATATATTTATCAAAATAAGTATGATGATCACCATTCTGTTGAACGATCTTTCCT	1506
QY	461	GlnTyrMetValProHisAsnLeuGIuaRThRhgIuaIaGlyMetLysCysLeuTyrTyrLeu	480
Db	1507	CAATRCATGTTCTCCACAAATTTAGAAACTACAGACCGATGAATGCTTATATATTCTTG	1566
QY	481	TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGlnMetTyrLysCysGlnAsn	500
Db	1567	TATGCCACACATGAGATTTAAATAGCTGTGAAGAGATTGAATGAATGTGAATGTCAAAAT	1628
QY	501	LeuLeuAtrGHisGlnValLysAspLeuAspLeuIleLysGlnProLysThrAspAla	520
Db	1627	CTGCCTCCACATCAAGTAAAGATTTCCTTACCTTGATTAAAGCAACCCAAACAGATCC	1686
QY	521	SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly	540
Db	1687	AGTGCAGAGCCATTTTTCAAAGTGTGTTATTACAGAAATTTACCTGATCTTGCT	1748
QY	541	LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg	560
Db	1747	AAGGCTCAGAGTTTCATGAGAATAATTCACACAGGTGTAGAAAGATGATGAGAAATAAGA	1806
QY	561	LysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnAlaGlnGlyCysVal	580
Db	1807	AAGCAGTTAGAGTCTTGTTAGTCCACATGCTCTCCCAAGCAGCGCTGAAGGTTGTGTG	1866
QY	581	ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGlnMet	600
Db	1867	CGTGAATTACTAAGAATGTGGGCAACCCCAACAGCCTACAAATCTTTCTGTGAATG	1926
QY	601	IleLysPheLeuLeuGlnArgIleAlaIeProValHisIleAspThrGlnSerIleSerAla	620
Db	1927	ATCAAGTTCTCTGTGGAGAGGTAGCACCTGTGCACATAGATACCAGATTCATACGTCT	1986
QY	621	LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlnGlyVal	640
Db	1987	CTTATTTAAACAAGTGCAACAAATCAATAGATGGACAGCAGATGTAGTAAGATGAGAGGTCTT	2046
QY	641	ProThrAspGlnAlaIleArgAlaGlyLeuGlnLeuLeuLysValLeuSerPheThrHis	660
Db	2047	CCAACTGTATCAAGCCATAGACAGCAGGTCTTGACCTTTAAGATCTCTCATTTTACAT	2106
QY	661	ProIleSerPheHisSerAlaGlnThrPheGlnSerLeuLeuAlaCysIleuLysMetAsp	680
Db	2107	CCCACTCATTTTCATTTCTGCTGGAACAATTGATCATATTACGCGCTGTCTGCAAAATGAT	2166
QY	681	AspGluLysValaIaGlnAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGln	700
Db	2167	GATGAAAAGAAGTAGCAGAGCTGCATCAAAATTTTCAAAAACACAGCAAGCAAAATTTGAA	2226
QY	701	GlnAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys	720
Db	2227	GAGGATTTTTCACACATCAGATCAGACCTTGTCTCTGTTTACATCAACAATCTTAAAAA	2286
QY	721	GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys	740
Db	2287	GGACCCCCCGTCAAGGCCAAATATATGCCATTATTTGTATATCCATGGCATTTTTCATAGTAA	2346
QY	741	GluThrGlnPheAlaGlnIlePheGlnProLeuHisLysSerLeuAspProSerAsnLeu	760

```

|||||
Db 2347 GAGACCCAGATTGGACACAGATATTGAGCCTCGCATTAAGAGCCAGATCCAGCAACCTG 2406
OY 761 GJUHLSleuIIeThProleuValThrlIeglyHISleAlaIleuAlaProaspGln 780
Db 2407 GACATCTTCATTAACACCATGGTTACTATGGTCAATATTGCTCTCTCCACCATATCA 2466
OY 781 PheAlaAlaProTrpLysSerTrpAlaIaThrPheIleValIlyAspLeuIleuIleAsn 800
Db 2467 TTTGCTGCTCCCTTGGAAATCTTTGGTAGCTACTTTTCATTGTGAAGATCTTCTCATGAT 2526
OY 801 AspaArgLeuProGlyLysLysThrThrlLysLeuTrpValProaspGluIuValSerPro 820
Db 2527 GATCGGCTCCAGGAGAAAAAGACAACCTTGGGTTCCAGATGAAGAACATATCTCT 2586
OY 821 GluTrpMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuIlyMetLys 840
Db 2587 GAGCAATGGTCATAAATTCAGGCTATTAAATGATGGTTCCGATGCTACTTGGAATGAAA 2646
OY 841 AsnaAsnHISerLysSerGlyThrSerThrlLeuArgLeuLeuThrThrlIleuHISer 860
Db 2647 AATATTCACAGTAATTCAGAACTCTACCTTAAGATTCTTAACAACATATTGCAATAGT 2706
OY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProaspMetSerArgLeuArgLeu 880
Db 2707 GATGAGAGACTTGACAGAACAGGAGAAATTAAGTAACAGATATGTCACGCTGAGACTT 2766
OY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProGlyThrlHISgluIleIleThr 900
Db 2767 GCTGTGGAGAGTGTATTTGTGAAGCTGGCACAAAGCCCTGTACCATGAAATCATCA 2826
OY 901 LeuGlnGlnTrpGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrlGlnValArgGln 920
Db 2827 TTAGAACAATATCACACTATGTGCATTAGCTATCAACGATGATGCTATCAATAAGACA 2886
OY 921 ValPheAlaGlnLysLeuHISLysGlyLeuSerArgLeuArgLeuProLeuGluTrpMet 940
Db 2887 GTGTTGGCCAGAAACCTTCACAAAGCCCTTCCGTTTACGGCTTCCACTTGAGTATATG 2946
OY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHISAlaArgGln 960
Db 2947 GCAATCTGTGCCCTTTGTGCACAAAGATCCCTGAAGAGAGAGAAAGCTCATCTAGGCCA 3006
OY 961 CysLeuValLysAsnIleAsnValArgArgGluTrpLysGlnHISAlaIleValSer 980
Db 3007 TGTGTGGTGAATAATATATAGTAAAGGGGAGTATCGAAGCAGACATGSCACTTTAGT 3066
OY 981 GluLysLeuLeuSerLeuLeuProGluTrpValValProTrpThrlIleHISleuAla 1000
Db 3067 GAAAAATTAATGCTCTCTACACAGATATGTTCCATATACAAATTCACCTTTGGCA 3126
OY 1001 HISAspProAspTrpValLysValGlnAspIleGlnGlnLeuLysAspValLysGluCys 1020
Db 3127 CATGACCCAGATATATGTCAAGATACAGATATGAACAACCTTAAGATGTTAAAGATGT 3186
OY 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHISAlaPheIle 1040
Db 3187 CTTTGGTTTGTCTGGAATATTAATGGCTAAATAATGAATAATACAGTCACGCTTTATC 3246
OY 1041 ArgLysMetValGluAsnIleLysGlnThrlLysAspAlaGlnGlyProaspAspAlaLys 1060
Db 3247 AGAAGAGTGTGAAGAAATATTAAACAACAAGATGGCCAGAGCAGCATATATCAAAA 3306
OY 1061 MetAsnGluLysLeuTrpThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3307 ATGAAATGAAAAACTTCACACTGTGTGATGCTGTCATGATATATCATATGCAAGAGT 3366
OY 1081 ThrThrTrpSerLeuGluSerProLysAspProValLeuProAlaArgPheMetThrlGln 1100
Db 3367 ACTACATACAGATTGGATCTCCCTTAAGACCCGAGTACTACAGCTCGTTTCTTCACTCA 3426
OY 1101 ProaspLysAsnPheSerAsnThrlLysAsnTrpLeuProProGluMetLysSerPhe 1120
|||||

```

```

Db 3427 CCTGACAGAAATTTAGTAACACCAAAAATTAATCTGCTCTGAAATGAATCATTTTTC 3486
OY 1121 ThrProGlyLysProLysThrlThrasnValIleuGlyAlaValAsnLysProIleuSer 1140
Db 3487 ACTCTGGAAAACCTAAACCAACATGTTCTAGAGGCTGTAAACAGCCACTTTCATCA 3546
OY 1141 AlaGlyLysGlnSerGlnThrlLysSerSerArgMetGluThrValSerAsnIleSer 1160
Db 3547 GCAGGCAGACATCTCAGACCAAAATCATACGAATGGAACACTGAAGCAATGCACAGC 3606
OY 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3607 ACCCTCAAAATCCAAAGCTCTCTCGAAGAAATAAAGGAGGCTGATATGCTCGAAATGAT 3666
OY 1181 HISSerGluAsnGluAspTrpThrlMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3667 CACAGTGAAGAAATGAAATTAACAAATGCTTCACTTGGCCGGGAAAAAAGTGCAAG 3726
OY 1201 ArgAspAspSerAspLeuValArgSerGluLeuGlyLysProArgGlyArgLysLysThr 1220
Db 3727 AGAGACAGACTGATCTTCTTAAGGCTGTGAATTGGAGAACCTTAAGAGCGAGAAAAAGC 3786
OY 1221 ProValThrlGluGlnGluLysLeuGlyMetAspAspLeuThrlLysLeuValGlnGlu 1240
Db 3787 CCCGTCACAGAACAGAGAGAGAAATTAAGTATGATGACTGACTTAAGTTGGTACAGCA 3846
OY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHISThrlAlaSerLysSerAsp 1260
Db 3847 CAGAAACCTAAAGGAGTCAAGCAAGTCCGAAAAAGGCAATACGCTTCAAAATCTGAT 3906
OY 1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
Db 3907 GAAACAGAGTGCGCTGAGAGAAAAAGAGCCCAAGAAGATATTGAAATAACATGAA 3966
OY 1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGly 1300
Db 3967 CAGAAATAGCCGCCCAAAAAGGTTAAAGAGGCCACCAACCAAACTTGTGTGAGGT 4026
OY 1301 ThrProLysGluGluProThrlMetLysThrlSerLysGlySerLysLysSerGly 1320
Db 4027 ACACCAAAAGAAAGACCAACAAATGAACTTTAAAAAAGAAAGCAAAAAAATCTGCA 4086
OY 1321 ProProAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrlGluGlnLys 1340
Db 4087 CCTCAGCAGCAGAGAGAGAGAGAGAAAGAAAGCAAAAGTGAAATTCGGAACAGAA 4146
OY 1341 SerLysSerLysGlnHISArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Db 4147 TCCAAAAGCCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGCAAGAAATCTCTGA 4206
OY 1361 SerSerAlaIleGluSerThrlGlnSerThrlProGlnLysGlyArgGlyArgProSerLys 1380
Db 4207 TCTAGTGCATTAATTCACACAGATCCACACAGTCCACACAGAAAGACGAGAAACATCAAAA 4266
OY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4267 ACGGCATCACCATCACACCAACCAAAAAAATATGTC 4299

```

RESULT 3
 HS095825 5253 bp mRNA linear PRI 05-APR-1999
 LOCUS Human androgen-induced prostate proliferative shutoff associated
 DEFINITION protein (AS3) mRNA, complete cds.
 ACCESSION U95825
 VERSION U95825.2 GI:4559409
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5253)
 Geck, P., Szelei, J., Jimenez, J., Soto, A.M. and Sonnenschein, C.
 Androgen-induced proliferative shutoff in prostate cancer cells

JOURNAL Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
REFERENCE 2 (bases 1 to 5253)
AUTHORS Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M.
TITLE Early gene expression during androgen-induced inhibition of proliferation of prostate cancer cells: a new suppressor candidate on chromosome 13, in the BRCA2-R1 locus
JOURNAL J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
MEDLINE 99229875
PUBMED 10215036
REFERENCE 3 (bases 1 to 5253)
AUTHORS Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REFERENCE 4 (bases 1 to 5253)
AUTHORS Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M.
TITLE Direct Submission
JOURNAL Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REMARK Sequence update by submitter
COMMENT On Apr 5, 1999 this sequence version replaced gi:4539617.
FEATURES
source
1..5253
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
/cell_line="LNCap"
/cell_type="cancer cells"
/tissue_type="prostate"
/note="exon 1 maps to p1 Artificial Chromosome PAC26H23; exons 2-24 map to cosmid 267p19; exons 25-34 map to p1 Artificial Chromosome PAC 49j10"
1..5253
/gene="AS3"
1..46
/gene="AS3"
/number=1
47..173
/gene="AS3"
/number=2
66..4241
/gene="AS3"
/codon_start=1
/product="androgen-induced prostate proliferative shutoff associated protein"
/protein_id="AAD22134.2"
/db_xref="GI:4559410"
/translation="MAHSKRTNDGKITPPGVKEISDKISKEEMVRLKMYVTEND MDQSEEEKELYNLAHLASDFLKHDPKDVLLVACCLADIFRIYAPAPYTSPPK LKDIPEITRQLGLEPTKSPFNRYFLLENTAWVSYNICPELEDSNLTFTQLYAT LFSVINNGHQVYHMHVMDSSIIICGDTVSBELDTVYLVLPARKINLKOAYDA KALKRQAQAIERYITNFQVLLGKTSISDJSEHVSIDLLELYNDSHLISVLPQ LEERLKSNDNEERLQVVKLLAKMGAKDSASQNKPLWCQYLGREFNDIHPRLCEY KFSASHCMNHPLDAKDLTEYLKYRSHDPEAIDHVDIVSVTAAKDILVNPDLNF VGERTLIDKRWVRKREAMGLAQIYKAYLASAGKDAKOIANKDKLHHYONSID DRLLEIRIFAYVYVPHNLLETTERMKCLYLYATLIDNAVALNEMKCOMLROVND LLDLIKOPKTDASVKAFESVWVITRSLPPRGAODPMKFTOVLDEDEKTROLEV LVSPTCSQKQAEVCVREITKRLGNKOPTNPFLEKTLERLAVHIDTSSISLKO VNSIDGTADDEDEGVPTDQIRAGLELNLVSTHPISFSAETFEFSLACLMDSE KVAEALQIFKNTGSKLEEDFPHIRSALEFHLHRSKSGPPROAKYAHCIIHAFSSK EAOFAQIEPLKSLDPSNLEHLITPLVTIGHTALLADPOFAPLKSLVNFETKDL MNDRLPGKTTKTLMWDEEYSPETMVKIOAKIMVRLGLGMKNHSGSSTLRILTIT LUSDDGTLTEQGIKSPDMRKLRLAAGSAIVKLAQEPCTHEIITLBOYQICALAINDE CIOVROVFAQLKHLGSLRKLPLEYMAICALCAKDPYKERRAHARQCLVKNINVRRY LKQHAAYSEKLLSLLEPYVVPYIHLAHPDVVYKODIQLDKVCKLFWVEILWA KNENNSHAFIRKVENIKQTKDAQGPDADANEKLYVCDVAMNIIMSKSTYSLESP KDVLVPARFETQDKNFSNTKNLPGMKSEFTPGKRTTNVLCGVNKPISAGKQSO TKSRMETVGNASSSSNPSRGRIGRLSDSEMDSHENEDYTMSPLPGRKSRDQD DIVRSELKPRGRKTPVTREREKLGMDDTKLVOBRPFGSQRSRKRGTATSSDQO OMPBEKRLKEDLLENEDEONSPPKGRGRPPPLVGGGTPEEPTMTKTSKGSKRGSP PPAPEEEEREQSGNTEOKSKSKOHVSRRAOORASSSSAIESSTOSPQKGRGP

SKTSPSPQPKKNV"
174..377
/gene="AS3"
/number=3
378..464
/gene="AS3"
/number=4
465..562
/gene="AS3"
/number=5
563..689
/gene="AS3"
/number=6
690..770
/gene="AS3"
/number=7
771..911
/gene="AS3"
/number=8
912..1027
/gene="AS3"
/number=9
1028..1122
/gene="AS3"
/number=10
1123..1268
/gene="AS3"
/number=11
1269..1420
/gene="AS3"
/number=12
1421..1534
/gene="AS3"
/number=13
1535..1616
/gene="AS3"
/number=14
1617..1665
/gene="AS3"
/number=15
1666..1805
/gene="AS3"
/number=16
1806..1921
/gene="AS3"
/number=17
1922..2027
/gene="AS3"
/number=18
2028..2188
/gene="AS3"
/number=19
2189..2312
/gene="AS3"
/number=20
2313..2471
/gene="AS3"
/number=21
2472..2540
/gene="AS3"
/number=22
2541..2677
/gene="AS3"
/number=23
2678..2801
/gene="AS3"
/number=24
2802..3006
/gene="AS3"
/number=25
3007..3121
/gene="AS3"
/number=26

exon 3122.. 3254
/gene="AS3"
/number=27
3255.. 3374
/gene="AS3"
/number=28
3375.. 3437
/gene="AS3"
/number=29
3438.. 3583
/gene="AS3"
/number=30
3584.. 3689
/gene="AS3"
/number=31
3690.. 4129
/gene="AS3"
/number=32
4130.. 4354
/gene="AS3"
/number=33
4355.. 4810
/gene="AS3"
exon
exon

BASE COUNT 1755 a 944 c 1074 g 1480 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 5253
Score: 7101.00 Matches: 1377
Percent Similarity: 99.28% Conservative: 4
Best Local Similarity: 98.99% Mismatches: 10
Query Match: 98.72% Indels: 0
Gaps: 0
DB: 9

US-09-512-581B-2 (1-1391) x HSU95825 (1-5253)

OY 1 MetLAHISserLYsThrArGThrAsnAspGlyLYsLIeThrTyrrProGlyValLYs 20
DB 66 ATGGCTCAATCAAGACTAGACCAATGATGGAATAATTCATATCCGCTGGGGTCAAG 125
OY 21 GluIleSerAspLYsLIeSerLYsGluGluMetValArgArgLeuLYsMetValLYs 40
DB 126 GAATATACAGATAAATATCTAAAGAGAGATGCGTGAACGATTAAAGATGGTTGTGAAA 185
OY 41 ThrPheMetAspMetAspGlnAspSerGluGluLYsGluLeuTYrLeuAsnLeuA 60
DB 186 ACTTTATGATATGACACGAGACTCTGAAGAGAAAGAGACTTTATTAAACCTAGCT 245
OY 61 LeuHISLeuAlaSerAspPhePheLeuLYsHISProGlyLYsAspValArgLeuLeuVal 80
DB 246 TTACATCTTGCTTCAGATTTTCTCAAGCATCTGTAAAGATGTCGCTTACTGTGA 305
OY 81 AlaCysCysLeuAlaAspLIePheArgLIeTYrAlaProGluAlaProTYrThrSerPro 100
DB 306 GCCGTGCTGCTTGGTGATATTTTCAAGATTTATGCTCTGAAGCTCCCTTACACATCCCT 365
OY 101 AspLYsLeuLYsAspLIePheMetPheLIeThrArgGlnLeuLYsGlyLeuGluAspThr 120
DB 366 GATAAACTAAAGAGATATTTTATGTTTAAACAAACAGTGAAGGGCTGAGAGATACA 425
OY 121 LysSerProGlnPheAsnArgTYrPheTYrLeuGluGluAsnLIeAlaTrpValLysSer 140
DB 426 AAGAGCCACAGTTCAATAGGATATTTTATTTGCTTGAGAACATTGGCTGCAAGTCA 485
OY 141 TyrAsnLIeCysPheGluLeuGluAspSerAsnGlnLIePheThrGlnLeuTYrArgThr 160
DB 486 TATAACATATGCTTTGAGTATAGAAATAGCAATGAAATTTTCACCCAGCTATACGAACC 545
OY 161 LeuPheSerValLIeAsnAsnGlyHISAsnGlnLYsValHISMetHISMetValAspLeu 180
DB 546 TTATTTTCAGTTATTAACAATAGCCACATATGCAATATGACATATGTTAGACCTT 605

OY 181 MetSerSerLIeLIeCysGluGlyAspThrValSerGlnGluLeuAspThrValLeu 200
DB 606 ATGAGCTTATTTATTTTGAAGAGTGATACAGTGTCTCGAGAGCTTTTGGATACGGTTTAA 665
OY 201 ValAsnLeuValProAlaHISLYsAsnLeuAsnLYsGlnAlaTYrAspLeuAlaLYsAla 220
DB 666 GTAAATCTGTACCTGCTCATAGAATTTAAACAAGCAAGCATATGATTTGGCAAGAGCT 725
OY 221 LeuLeuLYsArgThrAlaGlnAlaLIeGluProTYrLIeThrPhePheAsnGlnVal 240
DB 726 TTACTGAAGAGACAGCTCAAGCTATTTGAGCCATATATTACCAATTTTAAATCAGGTT 785
OY 241 LeuMetLeuGlyLYsThrSerLIeSerAspLeuSerGlnHISValPheAspLeuLIeLeu 260
DB 786 CTGATGCTTGGGAAAACATCTATCAGCATTTTGTCAAGCATGTCTCGACTTAAATTTTG 845
OY 261 GluLeuTYrAsnLIeAspSerHISLeuLeuLeuSerValLeuProGlnLeuGluPheLYs 280
DB 846 GAGCTTCATATATGATATGATATTCATTTGCTGCTCTCTGTATTACCCAGCTTGAATTTAAA 905
OY 281 LeuLYsSerAsnAspAsnGluGluArgLeuGlnValValLYsLeuLeuAlaLYsMetPhe 300
DB 906 TTAAAGAGCAATGATAATAGAGAGCGCTTACAGTTGTAAACTACTGGCGAAATGTTT 965
OY 301 GlyAlaLYsAspSerGluLeuAlaSerGlnAsnLYsProLeuTYrPGlnCysTYrLeuGly 320
DB 966 GGGCGAAAGATTTCAGATTTGGCTTCTCAAAAACAAGCCACTTTGGCACTGCTACTTGGGC 1025
OY 321 ArgPheAsnAspLIeHISValProLIeArgLeuGluCysValLYsPheAlaSerHISCys 340
DB 1026 AGGTTTATGATATTCATATGATATTCATATCCGCTGGAATGTGGAATTTGCTAGGCATGT 1085
OY 341 LeuMetAsnHISProAspLeuAlaLYsAspLeuThrGluTYrLeuLYsValArgSerHIS 360
DB 1086 CTCATGAAACCATCTCGATTTTGAAGAAACCTTAAACAGATGTCTTAAAGTGAAGTGCAT 1145
OY 361 AspProGluGluAlaLIeArgHISAspValLIeValSerLIeValThrAlaAlaLYsLYs 380
DB 1146 GACCTTGAGGAAGCTATTAGCATGATGTTTGTGTCAATATGATTACAGCTGCTTAAAG 1205
OY 381 AspLIeLeuLeuValAsnAspHISLeuLeuAsnPheValArgGluArgThrLeuAspLYs 400
DB 1206 GATATTTCTTGTGCAATATGATACACTTACTTAAATTTTGGGAGAGAAACATTAGACAAA 1265
OY 401 ArgTrpArgValArgLYsGluAlaMetMetGlyLeuAlaGlnLIeTYrLYsLYsTYrAla 420
DB 1266 CGATGGAGAGATACGCAAGAAAGACATGATGGAGACTTGGCCAAATTTATTAAGAAATATCT 1325
OY 421 LeuGlnSerAlaAlaGlyLYsAspAlaAlaLYsGlnLIeAlaTrpLIeLYsAspLYsLeu 440
DB 1326 TTACAGTACGACGCTGGAAAAGATGCTCCAAAACAGATAGCATGGATCAAAAGACAAATG 1385
OY 441 LeuHISLIeTYrTYrGlnAsnSerLIeAspAspArgLeuLeuValGluArgLIePheAla 460
DB 1386 CTACATATATATTAACAATAATAGTATGATGATGACACTACTTGTGAACGAGATCTTCT 1445
OY 461 GlnTYrMetValProHISAsnLeuGluTYrThrGluArgMetLYsCysLeuTYrTYrLeu 480
DB 1446 CAATPACAGGTTCCTCCACAATTTTAAACCTACAGAACGATGAATAATGCTTATTTACTTG 1505
OY 481 TyrAlaThrLeuAspLeuAsnAlaValLYsAlaLeuAsnGluMetTrpLYsCysGlnAsn 500
DB 1506 TATGCCACACTGGATTTTAAATGCTGTGAAGACATTGAATGAATGTGCAAAAT 1565
OY 501 LeuLeuArgHISGlnValLYsAspLeuLeuAspLeuLIeLYsGlnProLYsThrAspAla 520
DB 1566 CTGCTCCACATCAAGTAAGATTTGCTTGACTTGATTAAGCAACCAAAAAGATGCC 1625
OY 521 SerValLYsAlaLIePheSerLYsValMetValLIeThrArgAsnLeuProAspProGly 540
DB 1626 AGTGTCAAGGCCATATTTTCAAAAGCTGATGTTATTAACAAGATTTTACCGATCCCTGCT 1685
OY 541 LysAlaGlnAspPheMetLYsLYsPheThrGlnValLeuGluAspAspGluLYsLIeArg 560

```
|||||
Db 1686 AAGGCTCAGAGATTTCATGAAGAAATTCACACAGGCTTGAAGATGATGAGAAATATAGA 1745
QY 561 LysGluLeuGluValLeuValSerProThrCysSerCysGlnAlaGluGlyCysVal 580
Db 1746 AACCACTTAGAACTACTGTTAGTCCAAACATGCTCCGCAAGCAGGCTGAAGGTGGTG 1805
QY 581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
Db 1806 CGTGAATATACATAGAAGTTGGGCAACCCAAACAGCCTCAAAATCCTTCCCTCGAATATG 1865
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
Db 1866 ATCAAGTTTCTCTGGAGAGATAGACACCTGTCACATAGATACCGAATCTATCAGAGCT 1925
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
Db 1926 CTTATTAAACAATGAAACAATCAATGATGAGAACAGCATGATGAAAGATGAGGCTGTT 1985
QY 641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
Db 1986 CCAACGATCAACACCATCAGACAGGCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2045
QY 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuAlaCysLeuLysMetAsp 680
Db 2046 CCCATCTCATTTCTCTGCTGAACAATTTGAATTCATTTACTGGCCTGCTGAAATATGAT 2105
QY 681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2106 GATGAATAAATACAGACAGCTGCATCAAAATTTTCAAAACACAGAACCAAAATTTGAA 2165
QY 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLys 720
Db 2166 GAGCATTTTCCACACATCAGATCAGCCTTCTCTGCTTTTACATCAACAAATCTAAAAA 2225
QY 721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLys 740
Db 2226 GGACCCCCCGCTAAGCCAAATATGCCATTCATTTGATTCATCCGATATTTTCTAGAAA 2285
QY 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2286 GAGGCCAGTTTCACAGATATTTGAGCCCTCTGCATRAGACCTAGATCCACCAACCTG 2345
QY 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2346 GAACATCTCATACACCATGTTTACATATGTGCATATATCTCTCTTGCACCTGATCAA 2405
QY 781 PheAlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2406 TTGGCTGCTCTTGAATCTTTGGTAGCTACTTTCATTTGTGAAGATCTTTCATGAAAT 2465
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
Db 2466 GATCGGCTTCAGAGAAAAAGACAACTTGGGTTCCAGATGAGAAGATATCTCT 2525
QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
Db 2526 GAACACATGCTCAAAATTCAGGCTATTAAATGATGATGGTGCATGCTTGAATGAAA 2585
QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
Db 2586 AATAAATCACAGTAATCAGGAACCTTACTTAAGATGCTAACAAATAATTTGCAAGT 2645
QY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2646 GATGAGACTTCAGACAGACAGGAGGAAATTAATTAACACAGATATGTACACTTCAGACTT 2705
QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
Db 2706 GCTGCTGGAGTGTCTATTGTGAGCTGGCACAAGAACCCGTGTTACCATGAATATCATACA 2765
QY 901 LeuGluGlnTyrGlnLeuGlyAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
|||||
Db 2766 TTGAACAATATACAGTATGTGATTCATTCACGATGAATCATCAAGTAAACACA 2825
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
Db 2826 GTGTTGGCCAGAACTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAGTATAG 2885
QY 941 AlaIleCysAlaLeuGlyAlaLysAspProValLysGluArgAlaHisAlaArgGln 960
Db 2886 GCATCTGGGCCCTTTGTGCAAAAAGATCCGTAAAGAGAGAAAGACCTCATGTAGGCAA 2945
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaLysSer 980
Db 2946 TGTTTGGTGAATAATTAATGTAAAGCCGAGATATCTCAAGCAGCATCAGCTGTAAAT 3005
QY 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuAla 1000
Db 3006 GAAAAATTATGCTCTTCTTACCAGATATGTTGTTCCATATTCATTAACCTTTGGGA 3065
QY 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3066 CATGACCCAGATATATGCAAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGAAATG 3125
QY 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIle 1040
Db 3126 CTTTGGTTTCTCTGAAATATTAAATGCGTAAAAAAGAAATTAACATCAGCTTTTATC 3185
QY 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3186 AGAAATATGTTGAATAATATTAAAGCAAAACAAAGATCCCAAGACCATGTGCAAAA 3245
QY 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3246 ATGAATGAATAAACTGATACACTGTGTGATGTTGCCATGAATTCATCATGTCAAAGAT 3305
QY 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPheThrGln 1100
Db 3306 ACTACATACAGTTTGAATCTCTTAAAGACCGGTACTACACACTGCTTCTTCACTCAA 3365
QY 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
Db 3366 CTTGACAGAAATTTCAAGTAAACACAAATTAATCTCTCTCTGATGAATCAATTTTTC 3425
QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Db 3426 ACTCTCGGAAAACTTAAACAAACCAATGTTCTGTGAGCTGTGTAAACAGCCACTTTCATCA 3485
QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Db 3486 GCAGGCAAGCAATCTCAGACCAATCATCATCGATGGAACATGTAGCAATGCAAGCAGC 3545
QY 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3546 AGCTCAATTCACAGCTCTCTCTGGAAGAATTAAGAGGAGGCTGATGTTCTGAATAGAT 3605
QY 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3606 CACAGTGAATAATTAACATTAACAATGCTTTCACCTTTCGCGGGAAGAAAGATGCAAG 3665
QY 1201 ArgAspAspSerSerLeuValArgSerGluLeuGlnLysProArgGlyLysArgLysThr 1220
Db 3666 AGAGAGACTCTATCTTGTAAAGTCTGAAATTTGGAAGGCTTAAGAGCGAAGAAAAAAG 3725
QY 1221 ProValThrGluGlnGluGluLysLeuGlyMetCAspAspLeuThrLysLeuValGlnGlu 1240
Db 3726 CCCGTCACAGAAAGGAGGAGGAATTAAGTATGATGACTTGACTTAAGTTGGTACAGGAA 3785
QY 1241 GluLysProLysGlySerGlnArgSerArgLysArgGlyHisIleThrAlaSerGluSerAsp 1260
Db 3786 CAGAGACTTAAGAGGAGCTAGCGAAGTCCGAAAGAGGCGCATCGGCTTCAGATATCGAT 3845
QY 1261 GluGlnGlnThrProGluGluLysArgGluLysGluAspIleLeuGluLysGlnGluAspGlu 1280
Db 3846 GAACACAGTGGCTGAGGAAAGAGGCTCAAGAAAGATATATTAGAAAAATGAATGAA 3905
```

QY 1281 GlnaSerProProlYsGlyLysArgGlyArgProProlYsProLeuGlyGly 1300
|||||
Db 3906 CAGATAGTCCGCCCAAAAGGTAAGAGGCCACCAACCAACTCTTGTTGGAGGT 3965
QY 1301 ThrProlYsGluGluProThrMetYsThrSerYsGlySerYsLysSerGly 1320
|||||
Db 3966 ACACAAAAGAAAGACCAATGAAACTCTTAATAAAGAAAGCAAAAAAATCTGGA 4025
QY 1331 ProProAlaProGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
|||||
Db 4026 CCTCCAGACACAGAGAGAGAGAAAGAAAGAAAGGAAATACGAAACAGAG 4085
QY 1341 SerYsSerYsGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
|||||
Db 4086 TCACAAAGCAACAGACAGAGAGTGTCAAGAGAGACAGACAGAGCAATCTCTGAA 4145
QY 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnYsGlyArgGlyArgProSerYs 1380
|||||
Db 4146 TCTAGTGCATTTGAATCCACACAGTCCACACAGAAAGAGAGCAACATCAAAA 4205
QY 1381 ThrProSerProSerGlnProYsLysAsnVal 1391
|||||
Db 4206 ACGCATCATCATCAACCAAAAAAATGTG 4238
RESULT 4
AY102267 5317 bp mRNA linear ROD 01-JUN-2003
LOCUS
DEFINITION Mus musculus androgen-induced prostate proliferative shutoff
associated protein AS3 mRNA, complete cds.
ACCESSION
VERSION AY102267
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 5317)
Geck,P., Maffini,M., Sonnenschein,C. and Soto,A.M.
TITLE The AS3 proliferative arrest gene has an ancient eukaryotic
heritage and shows highly conserved functional domains in mice
Proc. Annu. Meet. Am. Assoc. Cancer Res. 43, 987 (2002)
JOURNAL
REFERENCE 2 (bases 1 to 5317)
AUTHORS Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
TITLE Extreme domain conservation and expression of the mouse AS3
proliferative arrest protein
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 5317)
AUTHORS Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
TITLE Direct Submission
Submitted (07-MAY-2002) Anatomy and Cell Biology, Tufts University
School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA
FEATURES
Source
1. 5317
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/C"
/db_xref="taxon:10090"
/chromosome="5"
/map="84cm; G2: between BRCA2 and KL"
/sex="male"
/tissue_type="brain"
147. 4487
/codon_start=1
/product="androgen-induced prostate proliferative shutoff
associated protein AS3"
/protein_id="AA052216.1"
/db_xref="GI:31321923"
/translation="MAHSKTRNDGKINYPGKVIKSPKISKEEMVRLLKVVTPFMD
MDPSEPEKEXYLNALILASDFLKHDPKRLVACCLADIFRITVAPAPYTPFM
LKDFMFTTROLGLEDYKSPQFNKFTFLKNIAWKNYCNFELDESENEFTQLYRT
LFSVINGNHGVHMYDLNLSIICEDDTVSQELDLVNLVPAHNLKQAVDLA
KALKRTAOAIEPYITNPNVLMIGKTSISDLSERHVDLILILEYINIDSHLLSVLPQ

LEFLKLSNDNEERLQVVKLLAKMEGAKDSBLASONKPLMOCYLCGRFNDIHVPIRECY
KASHCLNHNHBDLAKDLLEYLVKSHVSDHEDEIRHDIYVSIYTAKKDILHVNHLNF
VREHTICKRMVRKREKAMGLAOIKYKYSLSAARDKAOISWPKDKILHYVONSID
DRLVERIFAOYMPPHNLEETTERKCLLYLATLIDLANVAKLNEKRCOINLRIQVAD
LIDLKOPKTDASVKALFSKVAITRNLDPDGKODEKKKFTQVLEDEKTRKOLEL
VSPSCSKOAGCVREITKGNVROPNPRLNEMIKFLERIAPVHIDTESISALIKO
VNSIDGTADEDEGVPTDOAIRAGLETLAKLFTPTHPISFSAETFECLACIMDE
KVAEALIOIFKNTSKIEDEDPHRSALPLVHLHKSXSGPRROAKVYIHCIALIESK
ETFOAOFPEPHKSLDPSNLHILTPYVTIGHILALADPRAALPLSLVAFYIKDL
MNDRLPGKKTTRKLVYDPDEVSPEFTNIVKIQAIRKMYRRLGLKNNHSKSGSTLELLT
ILHSDGLTTEQKLSKPDMSRLRLAASVAVKLAODECYHEITILEDOQLCALINDE
CYOYROYFAOKLHKGSLRLPLEYMACALCAADPVERERARACOLVKNYTRER
LKOHASVSEKLSILPEYVYPTTHLALHDDPYKVOIDLEOLKDKVECLNVTLEIL
KNENNSHAFIRKMYENIKOTKODAGPDPTKNEKLYTCVDAAMNINISKSTYSLEP
KDPYLPAREFTPOPKNPSNTKNYLPEPEKSFPTGPKPTAVILAVNKLPSAGKOSQ
TKSSRMETVSNASSSNPSSGCRIRKGLRDSSEMHSSEDEDTMSPLBGRKSDKREDP
DLSELEKFRSKKAPVDPPEERLGHDDTLVQKREKSGRGRKGTASDBEQM
PEERHKEELLEENEDONSPPKRGKRPPLDGGTSGSEPPKTRKSKKKLVP
VVDDDEEEROIGMTEHKSROHRTSKRAOQARSPESTSVESSTOSTPOKGRPSK
APSPSOPPKIRVGRSKOVATKENDSSEMDVLQASSPVSDDTTOEGAEERDISVGNV
RRSSKRERRR"
BASE COUNT 1676 a 1100 c 1199 g 1342 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 5317
Score: 6942.00 Matches: 1344
Percent Similarity: 97.84% Conservative: 17
Best Local Similarity: 96.62% Mismatches: 28
Query Match: 96.51% Indels: 2
DB: 10 Gaps: 1
US-09-512-581b-2 (1-1391) x AY102267 (1-5317)
QY 1 MetalHisSerYsThrArgThrAsnAspGlyLysIleThrTyProProGlyValYs 20
Db 147 ATGGCTCATTCACAAAGACCAAGACCAACATGGGAAATTAACCTCCCTGGAGTCAAG 206
QY 21 GluIleSerAspYsIleSerYsGluGluMetValArgArgIleuYsMetValYs 40
Db 207 GAAATCTAGATTAATAATCTTAAAGAGAGATGGAGCGCGTTAAAGATGGTGTAA 266
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluGluYsGluLeuTyLeuAsnLeuAla 60
Db 267 ACTTTCATGAGCATGACGACGACGACTCTGAAGAGAAAGAACTTATCTTAACCTACT 326
QY 61 LeuHisLeuAlaSerAspPheLeuYsHisProGlyYsAspValArgLeuLeuVal 80
Db 327 TTACATCTTGGCTTGCATCTCTCCCAAGCATCCTGATTAAGATGTTCTGTTACTG 386
QY 81 AlaGlyCysLeuAlaAspIlePheArgIleTyArgAlaProGluAlaProTyThrSerPro 100
Db 387 GCTTGTCTGCTTGCATGATTTTTCAGGATTAATGCTCTGAGGCTCTTACACGTCCTCC 446
QY 101 AspYsLeuYsAspIlePheMetPheIleThrArgGlnLeuYsGlyLeuGluAspThr 120
Db 447 GATAAATCAAGAGATATTTATGTTTATTAACAAGGACACAAAGGAGCTGAAGATCA 506
QY 121 LysSerProGlnPheAsnArgTyPheTyLeuLeuGluAsnIleAlaTirValYsSer 140
Db 507 AAGAGCCCTCAGTTTAATAGATATTATTTACTTGAAGACATTCATGAGTTAAATCA 566
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyArgThr 160
Db 567 TATAACATATGCTTGAAGTGAAGATGCAATGAATCACTTACCACTTACAGCAACA 626
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnYsValHisMetHisMetValAspLeu 180
Db 627 TTATTTCTGATTAATAACAATAGCCACAAATGCAAGAGTTCAATATGCAATGCTGACCTC 686
QY 181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
Db 687 ATGAGTTCTATCATTTTGAAGTGATACCTGATCTCAGGAGCTCTTGAATACAGATTTAA 746

QY	201	ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaIleArgAspLeuAlaLysAla	220
Db	747	GTAATATCGTAGACCTGCCCATAGAACTTAACAAGCAAGCATATGATTTGGCAAGGCT	806
QY	221	LeuLeuLysArgThrAlaGlnAlaIleGluProGlyIleThrTrpPhePheAsnGlnVal	240
Db	807	TTTCGTGAAGAGACTGCTCAAGCTATTAACCATATATATACCATTATTTTAACCAAGGTT	866
QY	241	LeuMetLeuGlyLysThrSerLisSerAspLeuSerGlnHisValPheAspLeuIleLeu	260
Db	867	CTGATAGCTGGGAAACATCATATACAGTATTTGGTCTGAGCATGTGCTTGAATTTGATCTTG	926
QY	261	GluLeuGlyAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys	280
Db	927	GAACCTACCAACATTGACAGCCATTGTGTGTTCTGTCTGCCCCACCTTGAGATTGTAA	986
QY	281	LeuLysSerAsnAspAsnGlnGluIleArgLeuGlnValLysLeuLeuAlaLysMetPhe	300
Db	987	TTTAAAGACCATGATATAGAAAGAACGCCCTCCAGGTGTGTTAACTATGCGCAAAATGTT	1046
QY	301	GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuThrP6GlnCysTyrLeuGly	320
Db	1047	GGGGCAAAAGCATTCAGATTGGCTTCTCAGAAACAACCCCTCGCAGATGCTACTCGGCG	1106
QY	321	ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisLys	340
Db	1107	AGGTTTAATGATATCCATATGATACCAATACGCCCTGGAAATGTGAAATTTGCCAGCACATGC	1166
QY	341	LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis	360
Db	1167	CTCATAGAACCATCTCGATTATAGCAAGAATTAAACGAGTATCTTAAGTAGAGTACAT	1226
QY	361	AspProGluGlnAlaIleArgHisAspValIleValSerIleValThrAlaLysLys	380
Db	1227	GATCCCTGGAGAGCTATTAGACATGATCTTATTTGTCTAATAGAACGCTGTAAAAAG	1286
QY	381	AspIleLeuLeuValAsnAspHisIleLeuLeuAsnPheValArgGluArgThrLeuAspLys	400
Db	1287	GATATCTCTTCTTGCAATGATCACTCTACTCTCAATTTTGGACAGACCGCAACTTGGACAA	1346
QY	401	ArgTTPArgValArgLysGluAlaMetLeuGlyLeuAlaGlnIleTyrLysLysTyrAla	420
Db	1347	CGGTGAGATGACGTAAAGAACCATGATGGGACTGGCCAGATTTAATGAAGAAATCTCT	1406
QY	421	LeuGlnSerAlaIleGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu	440
Db	1407	TTACAGTCAGAGCTGGCAAGATGCTGCCAAACAGAAATCTCTGGCTCAAGCAACGCTG	1466
QY	441	LeuHisLisIleTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla	460
Db	1467	CTGCATATATACTATACAAACAGCATGCACACGCGCTACTTGTGGACGGATCTTTGCT	1526
QY	461	GlnTyrMetValProHisAsnLeuGlnIleThrThrGluArgMetLysCysLeuTyrTyrLeu	480
Db	1527	CAGTACATAGTCCCTCACATATTTGGAAACGACGGAACGATGAAGTCTTGATATCTTG	1586
QY	481	TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn	500
Db	1587	TACGCTACACGGAATCTTAATGCTGTGAACACATGAAATGGAATGCGAAATGTCAAAT	1646
QY	501	LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla	520
Db	1647	CTACTCCGACATCAAGTAAAGGATTTACTTGATTTAATTAACAACCAAAACAGATCC	1706
QY	521	SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly	540
Db	1707	AGTGCAGAGGCATATTTCTTAAGTGATGTATACACAAGAATTTCCAGATCTTGCT	1766
QY	541	LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg	560
Db	1767	AAGGCCCAAGACTTCATGAAAGATTCACACAAGTGTGGAAGATGATGAGAAAAATATACGG	1826

QY	561	LysGlnLeuGluValLeuValSerProThrcysSerCysLysGlnAlaGluGlyCysVal	580
Db	1827	AAACGCGCGAAGCCCTTTGATGTCACAACTGTTCTGCAAGCGCGTGAAGCGCTGTGTG	1886
QY	581	ArgGluIleThrLysLysLeuGlyAsnProLysGlnProIhrAsnProPhaLeuGlyMet	600
Db	1887	CGTGAATTAACCTAAGAACTTGGGCAACCCCTTAAGCAGCCTACAAACCCCTTTTGGAAATG	1946
QY	601	IleuysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla	620
Db	1947	ATCAAGTTTCTCTGGAGAGATAGCTCTGTACACATAGACACTGAATCATCATGTGCT	2006
QY	621	LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyAl	640
Db	2007	CTTATCAAGCAAGGAATTAATCAATGATGGAACAGCAGATGTATGAGATGAGAGGTGT	2066
QY	641	ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValIleSerPheThrHis	660
Db	2067	CCAACTGATCAAGCTATTCAGGGCAGCGCTTACGCTGCTGAAGGCGCTGTCTTTACGCAC	2126
QY	661	ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp	680
Db	2127	CCFATCTCCCTTCACTGCTGTGAGACGTTGATGCCCTTCTGCGCTGTCTGTAATGTGAT	2186
QY	681	AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu	700
Db	2187	GATGAGAAGGGGAGAGAGCTGCATCAAAATTTTCAAAACACAGGAGCAAGCAAAATGGA	2246
QY	701	GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLysLys	720
Db	2247	GAGATTTCCCTCACATCAGATTCAGCCCTGCTTCTGTACATCAATCAAAATCCAAAAA	2306
QY	721	GlyProProArgGlnAlaLysTyraIleHisCysIleHisAlaIlePheSerSerLys	740
Db	2307	GGACCACCTCGCGAGGCCCAATTAAGCCATTCATTGATTCATGCCATATTTTTCAGTAAA	2366
QY	741	GluThrGlnPheAlaGlnIlePheGlnProLeuHisLysSerLeuAspProSerAsnLeu	760
Db	2367	GAACCCCGTTGGACAGATATTTGAGCCTCTGATAAACCTTAGATCCAGCAAGCAACCTG	2426
QY	761	GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln	780
Db	2427	GAGCATCTTAATACCCCTGTGTGACATTTGTGTCATATGCTCTTTCGCACTGATCAG	2486
QY	781	PheAlaIaIaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn	800
Db	2487	TTCCGCTGCTCTGAAAGTCTTTGTGGTGCACACTTTCATTTGAGAGCACTCCATGATAT	2546
QY	801	AspArgLeuProGlyLysLysLysThrLysLeuTrpValProAspGluGluValSerPro	820
Db	2547	GACCGGCTTCAGAGAAAAAAGCAACATTAAGCTTTGGGTCCAGATGAGGAAGTCCACT	2606
QY	821	GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetCys	840
Db	2607	GAGCAATGTGCAAAATTCAGGCTTATAATGATGGTTCATGGCTACTTGGAAATGAG	2666
QY	841	AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer	860
Db	2667	AATATATCAGTAAGTGAAGAACTCCACCTCAGATTACTAACAACAGATACCTCATAGT	2726
QY	861	AspGluAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu	880
Db	2727	GATGGGGAATTTGACAGAAACAAGGCAAAATTTAGTAAACCGATATGCTCAGCGCTCAGACTT	2786
QY	881	AlaAlaIaGlySerAlaIleValLysLeuAlaGlnIleProCysTyuHisGluIleIleThr	900
Db	2787	GCTGTGGGAATGCTATTGTAGCTGAGCAAGGAGCGCTGTATCCAGAAATCATTTACA	2846
QY	901	LeuGluGlnLysLysLeuCysAlaLeuAlaIleAsnAspGluCysTyuGlnValArgGln	920
Db	2847	CTGAGAGCGATACAGACTGTGTGCATTAAGCCATCATATATGATGTATTAACAAGTCAAGCAG	2906
QY	921	ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnTyuMet	940


```

|||||
Db 2907 GTGTCGCCAGAACTTCACAAAGCCTTCCCGCTTACGGCTTCCCTTGAGTACATG 2966
Oy 941 AAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaArgGln 960
Db 2967 GCCATCTGTGCTCTTGTGGCAAGACCCTGTGAAGGAGAGCGCGCCATGCTGTAGACG 3026
Oy 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaValSer 980
Db 3027 TGCCTTGCGAAAGAACACACCTGTGAGGAGGAGTACCTGAAAGCAGCATGCAAGTGTAGT 3086
Oy 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3087 GAAAGATATATGTCTCTCTACAGAGATGTGTTTCATATACAAATTCACCTTTGGCA 3146
Oy 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3147 CATGACCCAGATTTATGTCAAGTACAGATATTTGAACAACCTTAAAGATGTGAAGAAATGCG 3206
Oy 1021 LeuTrrPheValIleGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Db 3207 CTTTGCTTGTCTGGAGATATTGATGCTTAAAGTGAACACACAGCCATGCAATTATTC 3266
Oy 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3267 AGAAAGATGTAAGAAATATTAAACAGACAAAGATGCTCAAGGACGAGATGATCAAAA 3326
Oy 1061 MetAsnGluLysLeuTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3327 ATGATGTGAAGATTTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGCTCCAAAGAC 3386
Oy 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheIleGln 1100
Db 3387 ACCACGTAACACCTGGAGTCTCTCAAGAGACCCCGGCGCCAGCTGGTTCACCCAG 3446
Oy 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
Db 3447 CCTGACAAGAAATTTAGTAACACCAAAAATTAACGCTCCAGAAATGAATCATTTTTC 3506
Oy 1121 ThrProGluLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Db 3507 ACTCCTGGAAACCTTAACCAAGCCAAATGTTCTCGAGACGATTAATAAACCTTTCATCA 3566
Oy 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Db 3567 GCAGGCAACAGTCTCAACCAATCATCAAGAAATGGAACCTGTGAGCAAGCAAGCAGC 3626
Oy 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3627 AGCTCCAAACCAACCTCTCTCGAAGGATCAAGGAGGAGCTTGATAGCTCGAATGATGAT 3686
Oy 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGluLysLysSerAspLys 1200
Db 3687 CACAGTGAAGATTAAGATTATACATGCTTCCCTTGGCCAGGAAAAAGTGCACAG 3746
Oy 1201 ArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArgGlyArgLysLysThr 1220
Db 3747 AGAGAAGACCTGTATCTT-----TCTGAGTTGAGAGAGAGCTTGAAGTGCAGAAAAAGCA 3800
Oy 1221 ProValThrGluGlnGluLysLeuGlnLysMetAspAspLeuThrLysLeuValGlnGln 1240
Db 3801 CCTGTCAACAGCCTCTCAAGAAATTAAGTATGATGACTAAGTAAGTGTGACAGAA 3860
Oy 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
Db 3861 CAGAAACCTTAAGCAGTACAGAGCGGCGGAAAAAGGCGCTACAGCTTACACTCAAC 3920
Oy 1261 GluGlnGlnTrrProGluGlnLysArgLeuLysGlnAspIleLeuGlnLysGlnLysAspGln 1280
Db 3921 GAGCAGCAGTGTGCTGAGGAGCAAGGAGCAAAAGAGAGCTCTGTGAAAAATGAGAGTACG 3980
Oy 1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGln 1300
|||||

```

```

Db 3981 CAGAACAGCCACCAAAAAAGGCGAAAAAGGAGCAGGCCACCAAAACCTTGTGTGGGGG 4040
Oy 1301 ThrProLysGluGlnProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
Db 4041 ACATCGAAGGAAAGAGCCCAATGAAATTCACAAAGAAAGCAACAAAAAAGCTTGTA 4100
Oy 1321 ProProAlaProGluGlnGluGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1340
Db 4101 CCTCTGTAGTAGACAGCAGATGAAAGACAGAAACAAATTTGAAACACAGAACATTAAG 4160
Oy 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Db 4161 TCAAAAGCAAAAGCAGCAGCCGACATCAAGAGAGCAGCAACAGAGCAGAAATCTCTGAA 4220
Oy 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db 4221 ACAAGTGCAGTTGATTCACACAGTCCACACAGCAAGAAAGACGAGAAACATCAAAA 4280
Oy 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4281 GCACCATCACCATCAACACCCCAAAAAAATC 4313

RESULT 5
LOCUS AK122414 5944 bp mRNA linear ROD 15-MAR-2003
DEFINITION Mus musculus mRNA for mKIAA0979 protein.
ACCESSION AK122414.1 GI:28972559
VERSION AK122414.1 GI:28972559
KEYWORDS FLI-CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yusa,S.,
Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
2 (bases 10, 35-48 (2003)
REFERENCE
Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
Direct Submission
Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- 3'-end
one pass sequencing.
FEATURES
source
Location/Qualifiers
1..5944
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbx00002"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
1..5944
/gene="mKIAA0979"
<1..3577
/gene="mKIAA0979"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=2
/evidence="not experimental"
/product="mKIAA0979 protein"
/protein_id="BAC55696.1"
/db_xref="GI:28972560"
/translational="FDLILEYNIIDSHLLSVLPQLEFKLSNDNEERLOVYKLLAKM

```

FGAKDSELASQNKPLWOCYLGRENDIHVIRLECVKFSASHCLNHNPLADLREYLVK
RSHDPEEAIHDIIVISIVTAAKDOIILVNHDLNFEVERLIDKRMVRKAMGLAOI
YKRSLOSASAGKDAKOISWVKDKLILHYONSIDRLVERIEFAQYVPHNETTER
MKCLYLYATLIDNAVKALNEMKCONLHROYVLDLIDLKOPDTASVAFISKVW
ITRNLDPGKAODFMKFTOVLEDEKTRLOEVALVSPCSQAEGCVREITKIGN
PKOPTNPFLEMIKFLERLAPVHIDTESIALIKOVAKSIDGTADDEDEVPDDAIR
AGLELVLSFTPHISFSAETESLACLKMDKDEKVAEALQIEKVTGSKIEDEPH
IRSLALPLVHNHRSKPKROAKYAHCIAHIFSSKEPOIEPPTKSLDPSLEHL
ITPLVTIGHIALALADQFAPLAKSLVATFYVKDILMDRLPGKTTIWLMPDEEVSPE
TMYKIOAKIMWHLGKNNHSGSTVRLTLTLLHSDRPLEGKTSKIPKMSRLV
LAAGSAIVKLADPECYEITLIDYOICALAIDECYOVQVFAOKLHKLSTRLEPL
EYMAICALKADPEKERRAHAROCIVANITVRRYLYQNAVSEKLSLLEPYVPT
IHLADPDYVKODIEQLKDVKECLMFVLEITLMAKNENSHATRMVENIKOTPA
OGPDUTMNEKLTVCVAMNIIIMSKSTYSLESDPVLPAFETQPDKNFSKTY
LPEPKSEFTPGKRTANVAVNKLPSAQSOTKSSMETVSNASSNSPSGR
IKGRILDSSEMDHSENEYTIMSSPLPGKKSREDPDLSELEKPSRKAAVYDEEKL
GMDULTLVEOKRPGKRGKRTASDEDOQWPEKHKELELNEDEONPSPK
GKGRPRKPLGGSGSKREPTMKTKGKNKKIYPPVYDDEDEEROGIMENKSKSO
HRTSKRAQOAESEPTSAVESTOSTPOKGRGRSPKAPSPQPRKIVGNSKOVAIKE
NDSSEMDVLOASSPVSDDTTQEGAEEDISVGNVRRRRSSKRERR"

BASE COUNT 1892 a 1151 c 1251 g 1650 t
ORIGIN

Alignment Scores:

Pred. No.: 1,84e-290 Length: 5944
Score: 5659.00 Matches: 1094
Percent Similarity: 97.71% Conservatave: 16
Best Local Similarity: 96.30% Mismatches: 24
Query Match: 78.67% Indels: 2
DB: 10 Gaps: 1

US-09-512-581b-2 (1-1391) x AK122414 (1-5944)

QY 256 PheaspleuileuGluLeuTYrAsnIleasPserHisLeuLeuLeuSerValleuPro 275
D 2 TTTCATTGGATCTTGGAACTGTACACATTTGCGCATTTTGGCTTCTGCTCCTGCC 61
QY 276 GluleuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGluValLysLeu 295
D 62 CAGCTTGAGTTAAATTAAAGAGCAATGATATGAGAACGCCCTCCAGGTGTTAAACTA 121
QY 296 LeuAlaLysMetPheGluAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyr 315
D 122 CTGGCGAAAGATTGGGGCAAAAGATTCAAGATTGGCTCTTCCGAAACAAACCCCTGG 181
QY 316 GlnCysTyrLeuGluArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335
D 182 CAGTGTACCTGGGACAGTTTAATGATATTCATTCACCAATACGCCCTGGATGTGTGAAA 241
QY 336 PheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeu 355
D 242 TTTCGCCAGCACTCCATCATGAAACCATTCGATTATTAACAAAGATTAAACAGATATCTT 301
QY 356 LysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIleVal 375
D 302 AAATGTAGGTCACATGATCTCGAAGAGCTATTAGACATGATGTATTATGTCTATAGTA 361
QY 376 ThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGlu 395
D 362 ACAGCTGCTAAAAAGATATTTCTTGTCTCAATGATACCTTACCATTTTGTGTAGAGAG 421
QY 396 ArgThrLeuAspLysArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle 415
D 422 CGAACAATTGGACAAACGGTGGAGAGTACGTAAAGAACCATGATGAGGACCTGGCCAGATT 481
QY 416 TyrTyrLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTyr 435
D 482 TATTAAGAAATACCTTTCACGTCCAGCTGGCAAAAGATGCTGCAAAAACAGATATCTGG 541
QY 436 IleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuVal 455
D 542 GTCAAGAGCAAGCTGCTGATATATCTATTCAGAACAGCATCGACGACCGGCTACTTGT 601

QY 456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
D 602 GAGGAGATCTTCTGCTCAGTACATAGGCCCTCAACAATTGGAAACGACGAAACGATGAG 661
QY 476 CysLeuTyrTyrLeuTyrAlaThrPheAspLeuAsnAlaValLysAlaLeuAsnGluMet 495
D 662 TGCCTGTATTACTTGTGACGCTACACGACTTCAATGCTGTGAAGCATTTGAATGAATG 721
QY 496 TyrPysCysGlnAsnLeuLeuAspGlnHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
D 722 TGGCAATGTCAAATCTACTCCGACATCAAGTAAAGATTTACTGTATTAATTAACAA 761
QY 516 ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsn 535
D 782 CCAAAAACAGATGCCAGCTGCAAGGCCATATTTTCTTAAATGATGTTATCACACAGAAAT 841
QY 536 LeuProAspProGluLysAlaGlnAspPheMetLysPheThrGlnValLysGluAsp 555
D 842 TTGCACATCTCGTTAAGGCCCAAGACTTCATGAGAAGTTCACACAGAGTGTGAGAGAT 901
QY 556 AspGluLysIleArgLysGlnLeuGluValLysValSerProThrCysSerCysLysGln 575
D 902 GATGAGAAATACGGAACACAGCTGGAAAGCCCTGTAGTCCACATGTTCTCTGCAAGCAG 961
QY 576 AlaGluGlyCysValArgGluIleThrLysLysLeuGluLysAspProLysGlnProThrAsn 595
D 962 GCTGAAGCGCTGTGCTGCTGAATTAACATTAAGTTGGGCAACCTTAAGCAGCTTACAAAC 1021
QY 596 ProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
D 1022 CTTTCTTGGAAATGATCAAGTTTCTCTTGAGAGGATACCTCTGTACACATAGACACT 1081
QY 616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp 635
D 1082 GAATCCATCAGTGTCTTATCAAGCAAGATGAATTAATCAATAGWGAACAGCAGATGAT 1141
QY 636 GluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuVal 655
D 1142 GAGATGAGGCTGTCCACATGATCAAGCATATCAGGCGACGCTTGAAGCTGCGAAGGTG 1201
QY 656 LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuAla 675
D 1202 CTGTCTTTACGACCTCATCTCTTCACTGTGCTAGACGTTGAGTCTCTTCTGCGCT 1261
QY 676 CysLeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
D 1262 TGCTGAAAGATGATGATGAGAAGGTGGCAAGGCTCCTACATAAATTTCAAAAACACA 1321
QY 696 GlySerLysIleGluLysAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
D 1322 GGAAGCAAAATTTGAAGAGGATTTCCCTCACATCAGATCAGCCTTGCTCTGTACTACAT 1381
QY 716 HisLysSerLysLysGluProProArgGlnAlaLysTyrAlaIleHisCysIleHisAla 735
D 1382 CACAATATCCAAAAAGAACCACTCGGACAGGCCAAATACCCATTCATTTATATTCAGCC 1441
QY 736 IlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeu 755
D 1442 ATATTTCTAGTAAAGAAACCAAGTTTGACACAGATATTAGGCTCGCATAAAGGCTA 1501
QY 756 AspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeu 775
D 1502 GATCCAAAGCAACCTGGAGCATCTTAATAACCCCTGTGCTACTATTTGCTCAATATGCTCT 1561
QY 776 LeuAlaProAspGlnPheAlaIleProTyrLysSerTyrAlaThrPheIleValLys 795
D 1562 CTGCACCTGATCACTTGCCTGCTCTGCAAGCTTTGGTGGCAACTTTCATATGTGAG 1621
QY 796 AspLeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuTyrValProAsp 815
D 1622 GACTCTCTCATGATGACGCGCTTCGAGAAAAAGCAACATAGCTTGGTGGTCCAGAT 1681
QY 816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyr 835

```

Db      1682 GAGGAAGTCCTCAGTCAGACATGGTCAAAATTCAGGCGATTAAATGATGTTGCGATGG 1741
Qy      836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThreuarGleuLeuThr 855
Db      1742 CTACTTGGAGTCAAAATATATCATCAGTAAGTCAGACTTCCACCCTCAGATTACTAACCA 1801
Qy      856 ThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMet 875
Db      1802 ACGATACGTCATAGTATGATGGGATTTGACAGAACAGGAAATTAATGTAACCAATATG 1861
Qy      876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGlnProCysThr 895
Db      1862 TCACCCCTGAGACTTGCTGCTGGAGTGCTATTGTGAAGCTGGACAGAGCCCTGTTAC 1921
Qy      896 HisGluIleIleThrLeuGlnGlnIleThrGlnLeuCysAlaLeuAlaIleAsnAspLys 915
Db      1922 CACGGAGATCATTTACACTGGAGCAGTACAGCTGTGTCATTTAGCCATCATGATGAGTGT 1981
Qy      916 TyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeu 935
Db      1982 TATCAAGTCAGCAGAGGTGTGCTCAGAACTTCACAAAGGCCCTTCCCGCTTACGGCTT 2041
Qy      936 ProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 955
Db      2042 CCCCTTGAGTACATGGCCATGCTGTGCTCTTGTGCCAAAGACCCTGTGAAGAGAGCGCA 2101
Qy      956 AlaHisAlaIleArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
Db      2102 GCCCATGTCTAGACATGCTTGGTGAAGAACATCACTGTGAGAGGAGACTGTAAGCAG 2161
Qy      976 HisAlaIleValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProThr 995
Db      2162 CATGACGCTGTTAGGAAAAATTTATGTCTCTCTTACACAGACTATGTTGTTCCATATACA 2221
Qy      996 IleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGlnGlnLeuLys 1015
Db      2222 ATTCACCTTTGGCAGATGACCCACAGATTATGCAAAATGACAGATATTGACCAACTTAAA 2281
Qy      1016 AspValLysGluCysLeuThrPheValLeuGlnIleLeuMetAlaLysAsnGlnAsn 1035
Db      2282 GATGGAGAAAGATGCTTGTGTGTCTGTGGAGATATTGATGCTTAAATAAGAAACAC 2341
Qy      1036 SerHisAlaPheIleArgLysMetValGlnAsnIleLysGlnThrLysAspAlaGlnGly 1055
Db      2342 AGCCATGCTATTATCAGAAAAATGCTAGAAAAATATTAACACAGCAAAAGATGCTTAAGA 2401
Qy      1056 ProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1075
Db      2402 CCAGATGATACAAAAATGAATGAATAATTTGTACACCGTGTGATGTTGCCATGAAACATC 2461
Qy      1076 IleMetSerLysSerThrThrTyrSerLeuGlnLysSerProLysAspProValLeuProAla 1095
Db      2462 ATCAATGTCACAGACACACGCTACGCTGTGAGTCTCCTAAGGACCCGCTGTGACACT 2521
Qy      1096 ArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGlu 1115
Db      2522 CGGTTTTTCACCCAGCCCTGACAAAGAAATTTAGTAACACCAAAAAATTTACTGCTTCAAA 2581
Qy      1116 MetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsn 1135
Db      2582 ATGAATATCATTTTTCACCTCTGAGAAACCTTAAACACAGCAATGTTCTCGAGACAGTTAAT 2641
Qy      1136 LysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal 1155
Db      2642 AAGCCACTTTCATACGACGCAAAACAGTCTCAGACCAATCATCAAGAAATGAAACTGTG 2701
Qy      1156 SerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAsp 1175
Db      2702 AGCAACGCAAGCAGAGCTCAACCCAGCTCTCTGTGAAGATCAAGGGAGGCTTAT 2761
Qy      1176 SerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlu 1195

```

```

Db      2762 AGCTCTGAATGATGATCACAGTCAAAATGAAATATATATACATGCTTTCACCTTGGCAGGA 2821
Qy      1136 LysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArg 1215
Db      2822 AAAAAAATGACAAAGAGAAAGACCCGTATCTT-----TCTGAGTTGGAGAAAGCTTGA 2875
Qy      1216 GlyArgLysLysThrProValThrGlnGlnGlnLysLysLeuGlyMetAspAspLeuThr 1235
Db      2876 AGTCGAAAAAAGCACCCTGTCTACAGACCCCTGAAGAAATTAATGATGATGACTTACT 2935
Qy      1236 LysLeuValGlnGlnGlnLysProLysGlySerGlnArgSerArgLysArgLysIleThr 1255
Db      2936 AAGTTGTGACAGAGACAGAAACCTTAAGCAGATCAGCAGGCGGAAAAAGAGCCGTACA 2995
Qy      1256 AlaSerGluSerAspGlnGlnGlnThrProGluGlnLysArgLeuLysGluAspIleLeu 1275
Db      2996 GCTTCAGACTCAGACAGCAGCAGCAGGCGCTGTGAGAGAAAGAGCACAAAGAGAGCTCTG 3055
Qy      1276 GluAsnGluAspGlnGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys 1295
Db      3056 GAAATATGAGATGACACAGACAGCCACCACCAAAAGGCCAAAGAGCAGCCACCAAAA 3115
Qy      1296 ProLeuGlyGlyGlyThrProLysGluGlnLysProThrMetLysThrSerLysLysSer 1315
Db      3116 CCTCTGGTGGGGGACATTCGAAGAGAGGCCACCAATGAAACATCCAAAGAAAGCAAC 3175
Qy      1316 LysLysLysSerGlyProProAlaProGluGlnGlnGlnGlnGlnGlnGlnGlnGln 1335
Db      3176 AAGAAAAAACTTGTACCTCTGTACTACAGCAGTGAAGAAAGAAAGCAAAATTTGGA 3235
Qy      1336 AsnThrGlnGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355
Db      3236 AACACAGAACATTAAGTCAAAAAAGCAAGCAGCCACATCAAGAGAGCAGCAACAGAGA 3295
Qy      1356 AlaGluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArg 1375
Db      3296 GCAGAAATCTCTGAAACCAAGTCAAGTCAATTCACACAGTCCACACACAGAAAGAGCA 3355
Qy      1376 GlyArgProSerLysThrProSerProSerGlnProLysAsnVal 1391
Db      3356 GGAAAGCATCAAAAGCACCATTCACATTCACACACCCCAAAAAAATC 3403

RESULN 6
LOCUS   AF294791                      6744 bp    mRNA    linear    PRI 24-JUL-2002
DEFINITION Homo sapiens SCC-112 (SCC-112) mRNA, complete cds.
ACCESSION AF294791
VERSION  AF294791.1 GI:21951801
KEYWORDS
SOURCE  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  Kumar,D., Patel,S., Whiteside,T.L. and Kasid,U.
  Identification and characterization of SCC-112, a novel cell cycle
  regulated gene in human cancer cells
  JOURNAL Unpublished
  REFERENCE 2 (bases 1 to 6744)
  AUTHORS   Kumar,D. and Kasid,U.
  DIRECT SUBMISSION
  JOURNAL   Submitted (09-AUG-2000) Radiation, Medicine, 3970 Reservoir Rd, NW,
  Washington, DC 20007, USA
FEATURES
  source
  1..6744
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /chromosome="4"
    1..6744
    /gene="SCC-112"
    232..4125
    /gene="SCC-112"
  CDS

```


Db 1720 GGAAGATGATGACATAGCAAGAAATTTGCTGACCCCGGAAAGACACAAATTTTGTG 1779
 QY LysLysPheThrGlnValLeuGlnLysAspArgLysIleArgLysGlnLeuGlnValLeu 566
 Db 1780 AAGAAATTTAACCAAGGTTCTCGCGATGTAGAAACCTTGGCTCAGTTGAGATTATTA 1839
 QY 567 ValSerProThrCysSerCysLysGlnAlaGlnGlyCysValArgGluIleThrLysLys 586
 Db 1840 ATTACCCACACTGTTCTTGCAACACAGCATATTGTGTGAGAGAAATACCCGGAAA 1899
 QY LeuGlnAsnProLysGlnProThrAsnProPheLeuGlnMetIleLysPheLeuGln 606
 Db 1900 CTGGCAAAATCCTAACCAACCAACAATCTTTCTAGAGATGCTGAATTTCTGTGGAA 1959
 QY ArgIleAlaProValHisIleAspThrGlnUserIleSerAlaLeuIleLysGlnValAsn 626
 Db 1960 AGAATCGCACCTGTGACATTTGATTCAGAGGCCATAGCCACTAGTGAATTTGATGAT 2019
 QY LysSerIleAspGlyThrAlaAspArgLysAspGlnGlyValProThrAspGlnAlaIle 646
 Db 2020 AAGTCATAGAGGGGACAGCAGATGATGAGAGAGGGGTGTAAGTCCAGATACAGCTATC 2079
 QY ArgAlaGlyLeuGlnLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
 Db 2080 CGTTCAGGACTTGAACCTTCTTAAGCTTCTGCTTTTACACATCTACCTCGTTCCACTCT 2139
 QY AlaGluThrPheGlnSerLeuLeuLacCysLeuLysMetAspAspGlnLysValAlaGln 686
 Db 2140 GCAGAGACATATGATTCCTTGTACAGTCCCTAAGAAATGGAGAGATGAGGATGAGAA 2199
 QY AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlnGlnAspPheProHisIle 706
 Db 2200 GCTGCTATTCAAATTTTAGAAATACAGTCCACAAATAGAAACAGACCTTCCACAGATA 2259
 QY ArgSerAlaLeuLeuProValLeuHisHisLysSerLysGlyProProArgGlnAla 726
 Db 2260 CGATGACCTTAATTCATTTTACATTTTACAAAAAGAGAGGAGCTCCACCAACAGCA 2319
 QY LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGlyThrGlnPheAlaGln 746
 Db 2320 AAACAGGCTGTGCACATGTATACAGCCCATATTACAAATTAAGAAAGTCCAGTTCACAG 2379
 QY IlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGlnHisLeuIleThrPro 766
 Db 2380 ATTTTGGACCCACTAGTAGAGCTGAATGCTGATGTGCCAGAACCACTTATTACTCCA 2439
 QY LeuValIleThrIleGlnHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLys 786
 Db 2440 TTAGTTTCAATGGGCCACATTTCTATGTTAGCACCAATCAGTTGCTTCCCAATGAAA 2499
 QY SerTyrValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
 Db 2500 TCTGTAGTAGCAAAATTTTATTGTGAAAGTCTGCTAATGAATGACAGGCAACAGGTGAA 2559
 QY LysThrThrLysLeuThrPValProAspGlnGlnValSerProGluThrMetValLysIle 826
 Db 2560 AAGATGGAAGAACTGTGTCTCCAGATGAGAGGTTCCCTCGAAGTACTACAAAGATA 2619
 QY GlnAlaIleLysMetMetValArgThrPheLeuGlnGlyMetLysAsnAsnHisSerLysSer 846
 Db 2620 CAGGCAATTTAACTCTGTGTAGGTGGCTGTGGTATGAAAAACACAGCTTAATCT 2679
 QY GlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGln 866
 Db 2680 GCCAATTACACCTTCGGTTATTATACAGGATGTTGGTTAGTGAAGGGTACCTGACAGAG 2739
 QY GlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIle 886
 Db 2740 CAAAGAGAGATCAGTAATTCGATATGTCTGCTGCGATTAGCTGCGTGTATGCCATA 2799
 QY ValLysLeuAlaGlnGlnProCysTyrHisGluIleIleThrLeuGlnGlnIleThrLeu 906
 Db 2800 ATGAAGCTTGTCTCAGGAACCTTGTACCATGAAATTTATACCCGACAGAACAGTTTCACTC 2859

QY CysAlaLeuAlaIleAsnAspGlnLysCysTyrGlnValArgGlnValPheAlaGlnLysLeu 926
 Db 2860 TGTGCACTTGTATTAAATGATGATGTTTACCAAGTAAGGACATATTTTGCTAGAACTG 2919
 QY HisLysGlyLeuSerArgLeuArgLeuProLeuGlnGlyTyrMetAlaIleCysAlaLeuCys 946
 Db 2920 CATTAAGGCACTGTGAAGTTACTGTCCCATTTGAGATATATGCGCATCTTTCCTTGCT 2979
 QY AlaLysAspProValLysGlnArgArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
 Db 2980 GCCAAAGATCCCTGTGAAGAGAGAGAGACACACGACACACAAATGTTTACTGAAAAATATC 3039
 QY AsnValArgArgGlnLysThrLysIleHisAlaIleValSerGlnLysLeuLeuSerLeu 986
 Db 3040 AGTATACCGAGGGAATTCATTTAAGCAAAATCTTATGCTACTAGCAAAATTTATTAACCTG 3099
 QY LeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVal 1006
 Db 3100 TTGCGTAATATGATGTTCCATACATGATTCACCTGCTAGGCCATGATCCAGATTTTACA 3159
 QY LysValGlnAspIleGlnGlnLeuLysAspValLysGlyLysLeuThrPheValLeuGln 1026
 Db 3160 AGATCACAAGATGTGATCAGCTCTGATATCAAAAGAGTGCCTATGCTCATGCTTGA 3219
 QY IleLeuMetAlaLysAsnGlnLysAsnSerHisAlaPheIleArgLysMetValGlnAsn 1046
 Db 3220 GTTTTAATGACAAAGATGAAACAAATAGCCATGCTTATATGAAAGATGGCAGAGAAC 3279
 QY IleLysGlnThrLysAspAlaGlnIProAspAspAlaLysMetAsnGlnLysLeuTyr 1066
 Db 3280 ATCAAGTTAATCAGAGATGCCAGCTCCAGTGTGAATTCACCAAGACAAAGAAACCTGAT 3339
 QY ThrValLysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGln 1086
 Db 3340 ACAGTATGTAGTGTGCTCTGCTGTATTAATTAAGTAAAGCTTTGTGCATGCATCAT 3399
 QY SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspAsnPheSer 1106
 Db 3400 TCACCAAGAGACCCAGTCTCTCCCATGAAATTTTATACACAACTGAAAGAGACTTCGT 3459
 QY AsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysProLys 1126
 Db 3460 AACGATTAAGATTAATTTACAGAAAGACAGACTTCTGTGTAAACAGGAAACCCAAAG 3519
 QY ThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGln 1146
 Db 3520 CCTGCTGAGATCTAGGTGCACTAATTAAGCTTTATACAGCAACGGAAGAAACCTAT 3579
 QY ThrLysSerSerArgMetGlnThrValSer-----AsnAlaSerSerSerSerPro 1164
 Db 3580 GTTAACACACTGCAGCTGACAGCTGAAAGCAATATTAATGTAAATTCAGAGCTGAACCT 3639
 QY SerSerProGlyArgIleLysGlyArg---LeuAspSerSerGlnMetAspHisSerGln 1183
 Db 3640 TCAACCGGAATTCATCAACAGACAGATTCAGAGGACAGACAGATGAGATGTGAA 3699
 QY AsnGlnAsp-----TyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg 1201
 Db 3700 AATGAAGAGAACCTGTGAGGATTTATTTCACTACACACCTGTAAAGAAATATTCAC----- 3753
 QY AspAspSerAspLeuValArgSerGlnLeuGlnLysProArgGlyArgLysLeuThrPro 1221
 Db 3754 -----CCAA 3756
 QY ValThrGlnGlnGlnGlnLysLeuGlnLysMetAspAspLeuThrLysLeuValGlnGln 1241
 Db 3757 GTAAGGAATTAAGAA-----ATTAATCTGTATCGAGCTAAC-----CAG 3795
 QY LysProLysGlySerGlnArgSerArgLysArgGlyHisIleThrAlaSerGlnSerAspLys 1261
 Db 3796 GGCACATCAGCAGTGCAGGAGGAAAGAAAGAAACAGTAAACAGCAGCAGCTGTGACAGAAAT 3855


```

Db      1023 GGAAGCAAAATTTGAAAGAGATTTTCACACATCATGATCAGCTTCTCTCGTTTACAT 1082
Qy      716 HistysSerLysLysGlyProProArg-GlnAlaLysTyrLalIleHisCysIleHisAl 735
Db      1083 CACAACATCTAAAGGACCCCGCCGCAAGCCAAAATCCCATCTTCTATCCAGCC 1142
Qy      735 aIlePheSerLysGlyThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
Db      1143 GATATTTTCTAGTAAAGAGACTCAGCTTTCACAGATATTGAGCCCTGCTGATAGAGCCT 1202
Qy      755 uasProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLe 775
Db      1203 AGATCCAAAGCAACCTGGAAACATCATCAACACATGGTACTATGCTCATATATGCTCT 1262
Qy      775 uLeuAlaProAspGlnPheAlaAlaIleProTyrLysSerTyrPalaIleThrPheIleVally 795
Db      1263 CCTTGACACCTGATCAATTTGCTGCTCTTGAAATCTTTGGTAGCTACTTTCATTCATGYGA 1322
Qy      795 sasPLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTyrPalaProAs 815
Db      1323 AGATCTTCTCATGATGATGATCGGCTTCCAGGAAAAAGCAACATAACTTTGGTTCCAGA 1382
Qy      815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyr 835
Db      1383 TGAAGAAGTATCTCCTGAGACAAATGTCAAAATTCAGGCTATTAAATGATGGTTCCATG 1442
Qy      835 pLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeu-ArgLeuLeuT 855
Db      1443 GCTCTATGGCATGTAAMTTAATTCACAGTATATCAGGATCTTCTCTTAAGSATTCGTAA 1502
Qy      855 hTThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspM 875
Db      1503 CAAACATATGCCATCGTATGCGGATTTGGAGAACAGGCGGAAT---TAGTAACCAAGTA 1559
Qy      875 eLsSerArgLeuArgLeuAlaIleGlySerAlaIleValLysLeuAlaGlnGluProCysT 895
Db      1560 TGTCACGCTGAGACTTCTCCTGAGAGTGGCTATTTAGACCTGGCACAGACCCGTTT 1619
Qy      895 yThIsgLulIleIleThrLeuGlnGlnTyrGlnLeuGlnCysAlaLeuAlaIleAsnAspGlu 915
Db      1620 ACCATGGAATCTCAGCAATAGAACATATCAGTATGCTTGTGCTATGCTATCAACGATGAAT 1679
Qy      915 ySTyrgInValArgGlnValPhe-AlaGlnLysLeuHisLysGlyLeuSerArgLeuArg 934
Db      1680 GCTATCAAGTAAAGCAAGTGTGTTGGCCAGAAACTTCGCAAAAGGCTTTCCCGTTGCGG 1739
Qy      935 LeuProLeuGluTyrMetAlaIleCysAlaLeuGlnCysAlaLysAspProValLysGluArg 954
Db      1740 CTTCCACTTGAAGATATGCGCAATCTGTGCCCTTTGTGCAAAAAGATCCGTAAAGAGAGA 1799
Qy      955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
Db      1800 AGAGCTCATGCTAGGCAATGTTGTGTAATAATATTAATGTAAGCGGAGAC 1850

```

```

MEDLINE 98403880
PUBMED 9734811
REFERENCE 2 (bases 1 to 5177)
AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail: cdna@info.kazusa.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)

FEATURES
Source
1..5177
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="Hj03994"
/sex="male"
/tissue_type="brain"
/clone_lib="pbluescriptII SK plus"
/dev_stage="adult"
1..5177
/gene="KIAA0648"
<1..2558
/gene="KIAA0648"
/codon_start=3
/product="KIAA0648 protein"
/protein_id="BAA31623.1"
/db_xref="GI:3327110"
/translation="LYYDYASLDPRNAKALNDENKCOMLRSHVRELDLHKOPTSEA
NCSAMFGLMTIARNLPDPRGAODPVKFFNOVLDDDEKLRQLELISPTSCROADI
CVREIARLAMPKOPTNPELEKVFLEKRIEYDLSAIEALVLMKMSIEGADDE
EGEVSPDAIRSGLELKVLSFTPTSPSHAEYTESLQCLMEEDKVAEAIQIFRN
TGHKIEPDLPIRSLIPIILHOKAKRGPRHAKAVHCHIAIFNKEVQLQIFEPS
RSINADVEPOLITPLVSLGHTSMILAPPOFAPMKSVANFTYKDLMLNDRSTGRNCK
LWSPDERVSPYVLAKVQAIKILVRLCCHKRNOKSKVANSRTLRLSAMLYSGCDITBOK
RISKSDMSRLRLAAGSALIMKLAQECHEIITTPQDLCALVIDECYQVQITRQKL
HKALVKLLPLEYMAIFALCAKDIPYKERARAROCLEKNISIREYIKONPMATEKL
SLIPEYVYPMIHLIADHPDFTRSQDVPDLRIKCLFMLEVLMTKNENSHAFMK
MAENIKLTRDOASDESCTNKEIYTCVDAVCIVNSKSAICNADSPKQPLVPMKFTQ
PERPCNDKSYISEETRYLLITGKPRKPGVIGAVNKPISATGRKRPYVSTGTGFSNI
NVNSLBNPSTNRSRBOSSSEAEVTSSENEENPVRIISVTVKNIIDPVKNEINSDOA
TQGNISDRGRKRVYTAGAENIQDKDEKVEDSGPPAPSKPRGRKRRKSSQGNATK
NDDNLSPINKRRKRAVGOESPGLGLENAKAPKLODLAKRAVAERQIDLR"
BASE COUNT 1690 a 931 c 1037 g 1519 t
ORIGIN
Alignment Scores:
Pred. No.: 5,266-134 Length: 5177
Score: 2704.50 Matches: 551
Percent Similarity: 74.07% Conservative: 123
Best Local Similarity: 60.55% Mismatches: 174
Query Match: 37.60% Indels: 63
DB: 9 Gaps: 15
US-09-512-581b-2 (1-1391) x AB014548 (1-5177)
Qy 477 LeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTyr 496
Db 3 TTATATTACTTATATAGCTAGTGTTCAGTCCAAATGCTGTAAAGGCTCTCAAGCAAAATGCGG 62
Qy 497 LysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnPro 516
Db 63 AAGTGTCAAGACATGCTTCGAGGACCATTTAGCGGACCATTTGATTTGCACACAGAGGCT 122
Qy 517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
Db 123 ACATCAGAGGCTTAATGTTCTGTCGCAATTTGGAAGAACATGATGACCATCAAGAAATATTT 182
Qy 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnLysAsp 556
Db 183 CTTGACCCCGGGAAGCACAAAGATTTTGTGAAGAAATTTAAACAGGTTCTGCGCGATCAT 242
Qy 557 GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576

```

JOURNAL

Db 243 GAGAACTTCGCTCAGTTGAGTATTATTAATTAAGCCCAACCTGCTTCGAAACAAGCA 302
 QY 577 GUGLGCysValArgIuIleThrLysLysLeuGlyAsnProLysGlnProThrAsnPro 596
 Db 303 GATATTGTGTGAGAAATAGCCGAAACCTTGCAATCTTAAGCAACAACAACCT 362
 QY 597 PheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGlu 616
 Db 363 TTTCGAGATGCTCAAAATTTCTGTGGAAAGAATGCGACCTGTCGACATTTGATTCGAA 422
 QY 617 SerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspArgIu 636
 Db 423 GCGATAGTGCATAGTGAATGAAATTAAGTCAATAGAGGGACAGAGATGATGAA 482
 QY 637 AspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuValLeu 656
 Db 483 GAGAGGGTGTAGTCCAGATACAGCTATCCGTTGAGACTTGACCTTTAAAGTTCTG 542
 QY 657 SerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
 Db 543 TCTTTACACATCTCACTGCTGTCACCTGACAGACATATAGTCTGTTACAGTGC 602
 QY 677 LeuLysMetAspArgIuLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly 696
 Db 603 CTAGAGATGAGATGACAGAGTACAGAGCTGCTATTCAAATTTTAAAGATACAGGT 662
 QY 697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHis 716
 Db 663 CACAATAATGAACAACAGCTTCCCGAGATTCGATCCAGCTTAATTCCTCATTTTACATCA 722
 QY 717 LysSerLysGlyProProArgGlnAlaLysThrAlaIleHisCysIleHisAlaIle 736
 Db 723 AAAGCAAGAGGGGTACTCCACACCAAGCAAAACAGCTGTGACCTGATACAGCCATA 782
 QY 737 PheSerSerLysGlnThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
 Db 783 TTACAAATTAAGATGACCTGTCGACAGATTTTGGACCTGACCTGATAGAGATCTGAAT 842
 QY 757 ProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuLeu 776
 Db 843 GGTGATGTCCAGAACACTTATTAACCTCATTTAGTTTCATTTGGGCCCATTTTCTGATGTA 902
 QY 777 AlaProAspGlnPheAlaAlaProThrLysSerThrValAlaThrPheIleValLysAsp 796
 Db 903 GCACCCAGATCAGTTGCTCCCAATGAATCTGTAGTACCAATTTTATTTGGAAGAT 962
 QY 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThrProAspGlu 816
 Db 963 CTCTTAATGATACAGGTGACAGGTGAAGAAAGATGAAAGTGTGCTCCAGATGAA 1022
 QY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeu 836
 Db 1023 GAGGTTTCCCTGAAAGTACGAAAGGTACAGGCAATTAACCTTCGTGTAAGTGTGCTG 1082
 QY 837 LeuGluMetLysAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 856
 Db 1083 TTGGGATGAAAAACAACAGCTTAATCTGCCAATTCACACCTTCGCTATTTATTCAGCG 1142
 QY 857 IleLeuHisSerAspArgLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
 Db 1143 ATGTGTGTTAGTAGGGGTGACCTGACAGAGCAAGCAAAAGGATCAAGTAATCTGATATCT 1202
 QY 877 ArgLeuArgPheAlaAlaGlySerAlaIleValLysLeuAlaGlnLysProCysThrHis 896
 Db 1203 CGCTTGCGATTACCTGCTGATGCCATATAGAGCTTCTCAGCAACCTTGTATACAT 1262
 QY 897 GluIleIleThrLeuGlnGlnIleThrGlnLysCysAlaLeuAlaIleAsnAspGlyCysThr 916
 Db 1263 GAAATATATACCCCAAGACAGTTTCAAGCTGTGCACTGTATTAATGATGAGTGTAC 1322
 QY 917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
 Db 1323 CAAGTAAGCAGATATTGCTCAGAAAGCTCATAAAGCAGCTTGGAAGTTACTGCTCCA 1382

QY 937 LeuGluThrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
 Db 1383 TTGGAGTATATGCGCATCTTGGCTTGTGTGGCAAAAGATCTCTGTGAAGAGAGAAAGCA 1442
 QY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgIuThrLeuLysGlnHis 976
 Db 1443 CACGACGACCAATGTTTACTGAAATAATATCAGATATCGCAGGAGAAATCAATTAAAGCAGAT 1502
 QY 977 AlaAlaValSerLysLeuLysLeuSerLeuLeuProGluThrValValProGlyThrIle 996
 Db 1503 CTTAGGCTTACAGAAATTTATATCAGCTGTGCTCGAATATGTACTTCCATATCATGATT 1562
 QY 997 HisLeuLeuAlaHisAspProAspThrValLysValGlnAspIleGluGlnLeuLysAsp 1016
 Db 1563 CACCTGCTAGCCCATGATCCAGATTTTACAGATTCACAAGATGTGATGACGTTCTGTGAT 1622
 QY 1017 ValLysGluCysLeuThrPheValIleGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
 Db 1623 ATCAAAAGAGGCTATGTTGTCATGCTTGAAGTTTATGCAAAAGATGAAACCAATAGC 1682
 QY 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
 Db 1683 CATGCCCTTATGAGAGAGATGCGCAGAGACATCAAGTTTACAGAGATGCCAGTCTCCA 1742
 QY 1057 AspAspAlaLysMetAsnGlnLysLeuLysLeuThrValCysAspValAlaMetAsnIleIle 1076
 Db 1743 GATGAATCCAGACAAATGAAAGATGAAAGCTGTATACAGTATGATGCTGCTGCTGTATTA 1802
 QY 1077 MetSerLysSerThrThrTySerLeuGluSerProLysAspProValLeuProAlaArg 1096
 Db 1803 AATAGTAAAGTCTTGTGTCATGTCAGATGCAACCAAGAGCCAGCTCCCAAGAAAGA 1862
 QY 1097 PhePheThrGlnProAspLysAsnProSerAspThrLysAspThrLeuProProGluMet 1116
 Db 1863 TTTTTCACACACTGAAAGAGCTTCTGTAAAGATAGAGGTATATTCACAGAGAGACA 1922
 QY 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGluAlaValAsnLys 1136
 Db 1923 AGAGTACTTGTGTTACAGAGAAAGCAAAAGCTGTGAGTACTAGTGCAGTAATTAAG 1982
 QY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
 Db 1983 CCTTTATCACAACAGGAGAAAGCAACCTATGTTAAGCACTGCGACCTGAGCTGGAAGC 2042
 QY 1157 AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg 1173
 Db 2043 AATATTATATTAATTCAGAGCTGCAAGCTTTCACCGGAATGATCAAGGAAACAGAGT 2102
 QY 1174 LeuAspSerSerGluMetAsnHisSerGluAsnGluAsp 1191
 Db 2103 TCAGAGGACAGCAAAAGTGTGAGTGAAGAAATGAAGAAACCTGTGAGAGATTATTTCA 2162
 QY 1192 ProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeu 1211
 Db 2163 GTCACACCTGTAAAGATATTATGAC 2186
 QY 1212 GluLysProArgLysArgLysLysThrProValThrGluGlnGluLysLeuGlyMet 1231
 Db 2187 2187 2213
 QY 1232 AspAspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLys 1251
 Db 2214 GATCAGAGCTAC 2258
 QY 1252 ArgGlyHisThrAlaSerGluSerAspGlu 1270
 Db 2259 AGAAGCAATACACAGCTGTGTCAGAAATATCCAAACAAAACAGATGAGAAA 2312
 QY 1271 LysGluAspIleLeuGluAsnGlnLysAspGluGlnAsnSerPro 1289
 Db 2313 2313

QY 1390 ArgGlyArgProProlysProleuGly---GlyGlyThrProLysGluGluProThrMet 1308
||||| ||||| ||| ||| |||
Db 2352 AAGAGCGTCGACCAAGCTGATCTGACATGCTACCAAAATGATGATCAAAAT 2411
QY 1309 LysThrSerLysGlySerLysLysSer---GlyProProlAlaProGluGlu 1327
||||| ||||| ||| ||| |||
Db 2412 AAACCTATTACAGGAGGAGAGAGAGCTGACGAGGCTCAGAGAGCCCTGGGCTTTG 2471
QY 1328 Glu-----GluGluGluArgGlnSerGlnSerLysnThrGlu 1338
||| ||| ||| ||| |||
Db 2472 GAACGACGATATGCCAAAGCACCACCAACTGCAAGATT-AGCCAAAAGGACAGCACCAGC 2530
QY 1339 GlnLys-----SerLysSerLysGlnHis-----ArgValSerArgArgAla 1352
::: ||| ||| ||| ||| |||
Db 2531 AGAAGACAAATTTGACTTACAAAGGTAATAATGATCATTTGCCAAAGGAGAAATGAAGGCC 2590
QY 1353 GlnGlnArgAlaGluSerProGluSer 1362
::: ||| ||| ||| ||| |||
Db 2591 AAACGAGAG---CAGGCTCCAGCTTCTGCA 2617
RESULT 9
BC039256
LOCUS BC039256 4252 bp mRNA linear PRI 06-NOV-2002
DEFINITION Homo sapiens, Similar to androgen-induced prostate proliferative
shutoff associated protein, clone MGC:33757 IMAGE:5295109, mRNA,
complete cds.
ACCESSION BC039256 GI:24657778
VERSION BC039256.1 GI:24657778
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4252)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-rt@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdc@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
FEATURES
SOURCE
1. 4252
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:33757 IMAGE:5295109"
/tissue_type="Testis"
/clone_lib="NIH-MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
71..1660
/codon_start=1
/product="Similar to androgen-induced prostate
proliferative shutoff associated protein"

/protein_id="AAH39256.1"
/db_xref="GI:24657779"
/translation="MAHSTRTNDKITYPGVKRISDKISKEMVRRLKMYKTFMD
MODSEEREKELYLNALHASPFLKHPKDVRLVACCLADIFRIYAEAPYTPDK
LKDIFMTITROLKLEDRKSPQFNRYFYLENIAMVKSYNICELEDSNEITPOLRT
LFSVINNGNORVHMHMDMSIICEGPTVSOELDVLVAVAHKMLNQAVDLA
KALKRTQALIPYITNFNOYLMGKTSISDSEHFPLILELNIOSHLLSVLPQ
LEFKLSNDEERLOVVKLAKMPKASELSAKSPKMGCTYGRNDINPRLCEV
KRASHCMNHPRLANDLREYLYKVRSHDPEALRHQYIVSYTAAKDLILVNDHLNF
VRETRLDKRWVRKEMKGLAOTIKRYALQSAAGDAQAMIKDKLHITYQNSID
DRLIVERIFAQYVMPVHNLNETTERMKCLYLYLATLIDLVNAKVVSNKFCSEHPLYDIGF
YKERTNFCILKCNLSVNIIV"
BASE COUNT 1407 a 664 c 792 g 1389 t
ORIGIN

Alignment Scores:
Pred. No.: 3.5e-124 Length: 4252
Score: 2517.00 Matches: 488
Percent Similarity: 98.80% Conservative: 4
Best Local Similarity: 97.99% Mismatches: 6
Query Match: 34.99% Indels: 0
DB: 9 Gaps: 0

US-09-512-581B-2 (1-1391) x BC039256 (1-4252)

QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
|||||
Db 71 ATGGCTCATTCAAAGCTAGGACCAATGAGAAAATTAATATCCGCTGGGCTCAAG 130
QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValLys 40
|||||
Db 131 GAAATATCGATGATAAATATCTAAAGAGAGATGTGAGACGATTAAGATGTTGGANA 190
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
|||||
Db 191 ACTTTATGATGATATGAGCAGCAGCTCTGAAGAAAGAACGACTTATTAACCTAGCT 250
QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
|||||
Db 251 TTACATCTGCTTCAGATTTTCTCAAGCATCCGATTAAGATGTTGCTTACTGCTGA 310
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
|||||
Db 311 GCCTGCTGCTGCTGATGATTTTTCAGGATTTATGCTCGAAGCTCCTTACACATCCCT 370
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGlnAspThr 120
|||||
Db 371 GATTAACCTAAAGATATATTTATTTAATCAAGACAGCTGAAGGCTAGAGATACA 430
QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTyrValLysSer 140
|||||
Db 431 AAGAGCCCAATTCATATGATATTTTATTTTACTTGACAAACATTTGCTGGCTCAAGCTA 490
QY 141 TyrAsnIleCysPheGluLeuGluLysPheAsnGluIlePheThrGlnLeuTyrArgThr 160
|||||
Db 491 TATAACATATGCTTTGAGTTAGATAGATACATTAATTTTCACACGCTATTAAGAACC 550
QY 161 LeuPheSerValIleAsnAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeu 180
|||||
Db 551 TTATTTTTCAGTTATTAACAAATGCGCCACATTCAGAAAGCTCATATGCTAGTACCTT 610
QY 181 MetSerSerIleIleCysGluGluLysPheValSerGlnGlnLeuLeuAspThrValLeu 200
|||||
Db 611 ATGACCTCTATTTATTTGTAAGCTATATACAGTGTCTCAGAGCTTTTGGATACGCTTTTA 670
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
|||||
Db 671 GTAATTCGTGATCGCTCATTAAGATTTAAACAAGCAAGCATATGATTTGGCAAGCT 730
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
|||||
Db 731 TTACGAGAGAGACAGCTCAACCTATTTGAGCATATTTATCAAAATTTTATATCAGGTT 790

```

OY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
    |||||
Db 791 CTGATGCTGGGAAACATCATCATCATGATGTCAGAGCATGCTTGAATATTTG 850
OY 261 GluLeuThrAsnIleAspSerHisLeuLeuSerValLeuProGlnLeuGlnPheLys 280
    |||||
Db 851 GAGCTCTACATATGATGATGATGATGCTGCTCTGCTTTTAAACCCAGCTGGAATTTAA 910
OY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
    |||||
Db 911 TTAAGAGCAATGATATGATGAGAGCGGCTACAACTGTTTAAACTAGCGCAAAATGTTT 970
OY 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGly 320
    |||||
Db 971 GGGGCAAGGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 1030
OY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
    |||||
Db 1031 AGGTTTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 1090
OY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
    |||||
Db 1091 CTGATGAAACATCCTGATTTAGCAAAAGACTTACAGAGTATCTTAAAGTGGCTCAT 1150
OY 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValIleThrAlaAlaLysLys 380
    |||||
Db 1151 GACCCCTAGAGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1210
OY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
    |||||
Db 1211 GATATTTCTTCTGGCAATGATCATCTTACTTATTTTGTGAGAGAGAGCACTTATGACAAA 1270
OY 401 ArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysTyrAla 420
    |||||
Db 1271 CGATGAGAGAGTACCAAGAAAGCATGATGGAGCTGCCCAATTTTAAAGAAATATGCT 1330
OY 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTyrIleLysAspLysLeu 440
    |||||
Db 1331 TTACAGTCAGAGCTGGAAGAAAGATGTCAGAAACAGATGATGATGATGATGATGATGATGAT 1390
OY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
    |||||
Db 1391 CTACATATATATATATCAAAATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1450
OY 461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
    |||||
Db 1451 CAATACATGATGCTCTCCTCACAATTTAGAAACTACAGAACGATGAATGCTTATTTACTTG 1510
OY 481 TyrIleThrLeuAspLeuAsnAlaValLysAlaLeuAsnGlnMetTyrPheLysCys 498
    |||||
Db 1511 TATGCCACACTGATTTAAATGCTGTAAGATGTTCAATATCAAAATTTCTGT 1564

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
CDS
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
US-09-512-581b-2 (1-1391) x AK098331 (1-2265)
OY 10 AspGlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGlu 29
    |||||
Db 182 GACCGGAAAGATCGCTTACCTCCGCGGCTAAAGAGATCAACGCAAGATCACACGAC 241
OY 30 GluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSer 49
    |||||
Db 242 GAGATGATCAAAACCCGTAAGATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 301
OY 50 GluGluGluLysGluLeuTyrTyrLeuAsnLeuAlaLeuHisIleLeuAlaSerAspPheLeu 69
    |||||
Db 302 GAACATGAAACACGACGATGATCCCATGACCTTGATGATGATGATGATGATGATGATGATGATGAT 361
OY 70 LysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
    |||||
Db 362 AGGAACCCCATTAAGATGTCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
OY 90 IleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPhe 109
    |||||
Db 422 ATCTATCCCGACGAGAGCTCCATATATCTCCATGATGATTAAGCATATATATTTGTTT 481

```

QY	110	llethrArgInleuLysGlyLeuGlnsphrIlysserProGlnPhasnArgtryThe	129
Db	482	ATTACCAACAACTTAAAGGTTTGGAGATACAAAGAGCTCCACAGACTTTAAATGATACATT	541
QY	130	TyrLeuLeuGlnAsnIleAlaTrpAllysserTrysnIleCySPheGluIdeuGlnAsp	149
Db	542	TATTATTATAGAGAAATTAGCTTGGTTAAATCATATAACATCTGCTTTGCAATTGGAAAT	601
QY	150	SerAsnGluIlePheThrGlnLeuTryArgThrIleuPheSerValIleAsnAsnGlyHis	169
Db	602	TGCATGTGAATTTTATTCACGCTTTTGGAACTCTCTTCAGTGATCAACAATAGCCAC	661
QY	170	AsnGlnIlyValIleMethIleMetValAspIleuMetserIleIleCyGlnGlnIlyAsp	189
Db	662	AATTAGAAAGTACAAATGACACATGCTGAATTTGATGAGCTTCATCATATGCAAGCTAT	721
QY	190	ThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsn	209
Db	722	GGAGTTCCTCAAGATTTATTGGACCTCCATCTTATTAACTCATATCTCGACATTAAGAC	781
QY	210	LeuAsnLysGlnAlaIaTryAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIle	229
Db	782	TTAAATAAACAGTCCCTTGGACCTTGCAAAAGTGCATTTGAAAGAACAGTCCAGACTT	841
QY	230	GluProTryIleThrThrPhePheAsnGlnValIleuMetIleGlyLysThrSerIleSer	249
Db	842	GAGCAGTCATGCTTCTGAATTTTTCATATCAAGTCCGTGGTGGGAGATCATCAGTAAGT	901
QY	250	AspLeuSerGlnHisValPheAspLeuIleLeuGlnLeuTryAsnIleAspSerHisLeu	269
Db	902	GATTGTGCAGAACATGTATTTGATCTGATTCAGCAACTTTTGGATATGATCCATCTTA	961
QY	270	LeuLeuSerValIleuProGlnLeuGlnPheLysLysLysSerAsnAspAsnGlnGlnArg	289
Db	962	TTATTATTCGTCATAGCCACACAGCTTGAAATTCAAAGACAAAGATGGAGAAAGCGA	1021
QY	290	LeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluIleuAlaSer	309
Db	1022	TTACCTGTGTTCACACTCTGTAGCTAAATTTGTTGGCTCCAAAGATTCGATTTGGCAC	1081
QY	310	GlnAsnLysProLeuIleTrpGlnCysTryIleuGlnArgPheAsnAspIleHisValProIle	329
Db	1082	CAGATGCTCCCTTTGGCAATGTGTTTGGACGATTAAATGATTCATGTGCTCTGTG	1141
QY	330	ArgLeuGlnCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys	349
Db	1142	AGATTAGAAAGTGAATTTGCCACATCATTTGTTAAATGAATTCACCCAGATTTAGCGAAG	1201
QY	350	AspLeuThrGluTryLeuLysValArgSerHisAspProGlnGluAlaIleArgHisAsp	369
Db	1202	GATGTCACAGAAATTTTAAAGGTAGATCACATGTCACAGAAAGAGCTATTCGTCAATAT	1261
QY	370	ValIleValSerIleValIleThrAlaAlaLysLysLysIleLeuLeuValAsnAspHisLeu	389
Db	1262	GTCATTCTTACTATATAACACGCTGCCAAGAGGACCTGGCCCTTAGTAAATGATCAGCTG	1321
QY	390	LeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMet	409
Db	1322	CTTGCTTTTGAAGGGAAGAACCTGCGATAAAGCGTGGCCAGTAAAGAAAGAGCTATG	1381
QY	410	MetCysLeuAlaGlnIleTryLysLysTryAlaLeuGlnInsAlaIleGlyLysAspAla	429
Db	1382	ATGGGCTGTGGCTACGCTTTATTAATAAATACGTCTTCATGGTGAACGAGGAAGGAACT	1441
QY	430	AlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTryTyrGlnAsnSerIle	449
Db	1442	GCAGAGAAAGACAGCTGGATTAAGACAAACTTCGATATTATTATACAGAAACGACTT	1501
QY	450	AspAspArgLeuLeuValGluArgIlePheAlaGlnIleuMetValProHisAsnLeuGlu	469
Db	1502	GAGCAACAAACGTGTGGTAGAAGAAAATCTTCTCAGTATCTTGTGCCCCACAAACCTGGA	1561

QY	470	ThrhgUaAgmMeLysCysLeuTyrTyrLeuYrAlaThrLeuAspLeuAlaVal	489
Db	1562	ACAGAGAGCAAAATGAATGCTTATATTACTTATATGCTACTTTGGATCCAAATGCGTGA	1621
QY	490	LysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu	509
Db	1622	AAAGCTGTCAACGAATAATGTGGAAGTGTGAGAAACATGCTTCCGAGCAGATGTCACGGAAC	1681
QY	510	LeuAspLeuLeuLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal	529
Db	1682	TTTGATTTGGCAACAGCACCCATCATCAAGGCTTAACCTTGCCATGTTTGGAAACATG	1741
QY	530	MetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe	549
Db	1742	ATGACACATGCAAAAGAAATTTGCTGCTGACCCCGGGAAGACACAAAGATTTGTGACAGAAAT	1801
QY	550	ThrGlnValLeuGlnAspAspGluLysIleArgLysGlnLeuGlnValLeuValSerPro	569
Db	1802	AACCAAGTTTCTGGCGCATGATGAGAAACATTCGTGCTCGATTGTGATTAATTAAGCCCA	1861
QY	570	ThrCysSerCysLysGlnAlaGlnGlyCysVal	580
Db	1862	ACGTGTTCTTGCAAAACAGACAGATATTTGTGTG	1894
RESULT 11			
LOCUS	BC041361	2705 bp	mRNA
DEFINITION	Homo sapiens, Similar to KIA00648 protein, clone MGC:43818		
ACCESSION	IMAGE:5273075, mRNA, complete cds.		
VERSION	BC041361		
KEYWORDS	BC041361.1 GI:27552793		
SOURCE	MGC.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 2705)		
JOURNAL	Strausberg,R. Direct Submission Submitted (16-DEC-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cga@db.remail.nih.gov Tissue Procurement: Mirkos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shliraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chiu, Readman Chiu, Chris Flell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schell, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 75 Row: c Column: 19. Location/Qualifiers 1..2705 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:43818 IMAGE:5273075" /tissue_type="Testis" /clone_id="NIH_MGC_97"		

CDS

/lab_host="DH10b"
/note="Vector: pBluescript"
545..2347
/codon_start=1
/product="Similar to KIAA0648 protein"
/protein_id="AAH1361.1"
/db_xref="GI:27552794"
/translation="MDPTAQPKPATALGVASADGKIAPGVKEIKDITPDEMIR
LKAVVTFEMDDODEKQOYLPLALHLSFEFLRPNKDVLLVACCLADIERIA
PEAPYSHDKIDPLFIETRIOLKLEPTKSPQPNRYEYLELAAWKSVCLEPDC
NEIFQIFRTLEFSVINSHKKYQMHMLDMSIIMEGDVTQELDSILINLPARK
NLNKQSDFLAKVLKRTVQTEACIANFNOVLVIGSSVDSSEHVEDLIQELFAD
PHLISVMPQLEFLKLSNDSEERLAVRLAKLFSGSDLATONRELMQCFREPD
IHVPRLSEYKFPASHCLMNHPRDLAKDLEVLKVSHPBEAIRDVYVITTAARDL
ALVNDOLGFEVERETLDKRWVRKEAMGAOLYKKYCLHGEAGKEAEKVSITKDL
LHYVONSIDDKLVEKTFNOYLPVPHLFEPEBKCYLYXASLDPRAYVALNEMKC
QNNLRSHVRLELDLHKQPTSEANCSANFGLMTAKLPLDPGKAODEVKFNOVLGSD
EKLRSQLELLISPSCKQADICVSKSYTFLFL"

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 3.7e-122 Length: 2705
Score: 2474.00 Matches: 460
Percent Similarity: 91.77% Conservative: 64
Best Local Similarity: 80.56% Mismatches: 47
Query Match: 34.39% Indels: 0
DB: Gaps: 9

US-09-512-581b-2 (1-1391) x BC041361 (1-2705)

QY 10 AspGlyLysIleThrTyrProGlyValAlaLysGluLeuSerAspLysIleSerLysGlu 29
DB 602 GAGGGGAGAGATCGCTTACCTCCGGGGTAAAGAGATCAACGACAGATCACACGAGAC 661
QY 30 GluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSer 49
DB 662 GAGATGATCAAAACCCCTGAAGATGCTGTGTAAGAAACCTTATGATGATGATGATGATGAT 721
QY 50 GluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisIleuAlaSerAspPhePheLeu 69
DB 722 GAAGATATAAAGACGCGATCTCCCTAGCCCTGATGATGATGATGATGATGATGATGATGAT 781
QY 70 LysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
DB 782 AGGAACCCCAATTAAGATGCGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 841
QY 90 IleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetCpHe 109
DB 842 ATCTATGCCCCAGAAAGCTCATATACCTTCCCATGATTAAGCATATATTTGCTTT 901
QY 110 IleThrArgGlnLeuLysGluLeuGluAspThrLysSerProGlnPheAsnArgTyrPhe 129
DB 902 ATTAACCGACAAATTAAGAGTTTGAGAGATACAAAGAGTCCACAGTTTAATAGATCTTT 961
QY 130 TyrLeuLeuGluAsnIleAlaTyrValLysSerTyrAsnIleCysPheGluLeuGluAsp 149
DB 962 TATTATTAGAGATTAGCTTGGTTAAATCATATACATCTGCTTGAATTGAGAGAT 1021
QY 150 SerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHis 169
DB 1022 TGCATAGAAATTTTATTCAGCTTTTGAACCTCTCTTCAGAGATCAACAAATAGGCAC 1081
QY 170 AsnGlnValHisMetHisMetValAspLeuMetSerIleIleCysGluGlyAsp 189
DB 1082 AATTAAGAGAGTACAAATGACATGCTAGATTTGATGAGCTTCTATCATCATGAGAGGCTGAT 1141
QY 190 ThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsn 209
DB 1142 GGAATTAATCAAGATATTTGAGACTCATTTCTATTAACCATCTCTGACATTAAGAAC 1201
QY 210 LeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIle 229

DB 1202 TTAATAAAGAGTCTTTGACCTTGCAAAAAGTCTATTGAAGAACAGTCCAGCATATT 1261
QY 230 GluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249
DB 1262 GAGGATGATGATGCTTAATTTTTCATCAAGAGTCTGCTGAGGAAGATCATGATGATGAT 1321
QY 250 AspLeuSerGlnHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu 269
DB 1322 GATTTCGACAAACATGATATTGATTCATTCAGGAAGCTTTTGTCTATAGATCCCATTTA 1381
QY 270 LeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluArg 289
DB 1382 TTATTATCCCTCATGCCACAGCTTGAATTCAACTAAAGAGCATATGTCAGAGAGCGGA 1441
QY 290 LeuGlnValValLysLeuLeuAlaLysMetPheGlyLalaLysAspSerGluLeuAlaSer 309
DB 1442 TTAGCTTTGTTGACACTTTCAGCTTAATTTGTTGGCTCCAAAGATTTGATTTGGACACA 1501
QY 310 GlnAsnLysProLeuTyrGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIle 329
DB 1502 CAGATGCTCTCTTGGCAATGTTTCTTGGAGGATTTATGATATTCATGCTTCCTGTG 1561
QY 330 ArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys 349
DB 1562 AGATTAGAAAGTGAATTTGGCAGCATGTTTATGATGATGATGATGATGATGATGATGATGAT 1621
QY 350 AspLeuThrGluTyrLeuLysValArgSerHisAspProGluGlnAlaIleArgHisAsp 369
DB 1622 GATCTCACAGAAATTTTAAAGGTTAGATCAATGATTCACAGAGATATTCCTCATGAT 1681
QY 370 ValIleValSerIleValThrAlaAlaLysAspIleLeuLeuValAsnAspHisLeu 389
DB 1682 GTCATGTTCTTAATAAAGAGTCCCAAGAGGAGCTGCTTATGATTAATGATGATGATGATGAT 1741
QY 390 LeuAsnPheValArgGluArgThrLeuAspLysArgTyrArgValArgLysGluAlaMet 409
DB 1742 CTGGGCTTTGTAAAGGAAGAACATGATGATGATGATGATGATGATGATGATGATGATGAT 1801
QY 410 MetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaIleArgLysAspAla 429
DB 1802 ATGGGTGCTGCTAGCTTTTAAAGAAATACCTGCTTCTATGATGATGATGATGATGATGATGAT 1861
QY 430 AlaLysGlnIleAlaTyrPheLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
DB 1862 GCAGAGAAAGTCAAGCTGATTAAGAGAACAACTCTGATATTATATCATGAGACAGCAT 1921
QY 450 AspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGlu 469
DB 1922 GACGACAACTGTTGTAAGAAATCTTGTGCTCAGATATCTGTCCCAACACTGGAA 1981
QY 470 ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaVal 489
DB 1982 ACAGAGAGAGAAATGAATCTTATATATTCTTATATGCTAGTTTGGATCCAAATGCTGTA 2041
QY 490 LysAlaLeuAsnGluMetTyrLysCysGlnAsnLeuArgHisGlnValLysAspLeu 509
DB 2042 AAAGCTCTCAACGAATGTGAAGTGCACAGACATGCTTGGAGCCATGATGACCAACTA 2101
QY 510 LeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal 529
DB 2102 TTGGATTGTGCAAGGCGCTCATCATCAGAGCTTAACCTGTTCCCAATTTGGAAGAACT 2161
QY 530 MetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 549
DB 2162 ATGACCTATACAAAGAAATTTGCTGACCCCGGGAAGACACAAATTTTGGAAGAAATTT 2221
QY 550 ThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGlnValLeuValSerPro 569
DB 2222 AACGAGTCTCGCGATGATGAGAAACTCGGCTCAGTGTGAGATTATTAATTAAGCCCA 2281
QY 570 ThrCysSerCysLysGlnAlaGluGlyCysVal 580
DB 2282 ACCGTCTTGTGCAAAACAGATATTTGTGTG 2314

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEMLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT	FEATURES	SOURCE		
BC021408	Mus musculus RIKEN cDNA 9030416H16 gene, mRNA (cDNA clone MGC:29439 IMAGE:3964501), complete cds.	BC021408	1	GI:18204099	MGC.	Mus musculus (house mouse)	Mus musculus	Enxarvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	1 (bases 1 to 1976)	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shamen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Cenciari, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McEran, K.J., Malek, J.A., Guarnata, P.H., Richards, S., Mowley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huyl, S.W., Villallon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Fahy, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, J.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	22388257	12477932	2 (bases 1 to 1976)	Strausberg, R.	Direct Submission	Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland. Web site: http://www.nisc.nih.gov/ Contact: ayce.mcgen@rl.nih.gov Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hachigigl, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Mastello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., Mcowall, J.L., Pearson, R., Skattilpob, S., Thomas, P.J., Touchman, J.W., Tsurgoun, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 39 Row: b Column: 24 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.	location/Qualifiers	1. .1976	/organism="Mus musculus"

gene	/mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone_MG:29439 IMAGE:3964501" /tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy." /clone_lib="NCI CGAP_Mam1" /lab_host="DH10B" /note="vector: pcMV-SPORT6" 1..1976 /gene="9030416H16Rik" /db_xref="locusID:71521" /db_xref="MG:1.918771" 130..1884 /codon_start=1 /product="Riken cDNA 9030416H16" /protein_id="AAH21408.1" /db_xref="GI:18204100" /db_xref="locusID:71521" /translation="MFGKLMITAKNLDPBGRKADQPVKKFNVOVGDEKRSOLELLI5 PTGCKADQVCVREIARKLANPKQPTNPFLEWVFLERIPVHIDSAISALAKLKM KISIGTADDEEGVSPDSAIRSGLELLKVFPTSHSSEYTESLQCLRMEDKX AELAIQIFRNTGHIETDLPQIRSTLIPILQKAKRGPHQAKAVHC IHAIFSNKEV QLAOIFPEPLSLNADVEQILITPLVSGHISMLAPOFASPMKSVYANFVTKDLNM DRSGEKNGKLMSPDEEYSPVLAIKVQAIKLVMILGKMGKNSKANSSTRILSAM VSEBDLTQKRIRKSDMSRLRLAASAMKAOEPCYHEITTPPOFLCAVINDECI QVROIIFAKLHKALVKLPLLEYMATIFALCAKDVKEERRAHROCLINKISIRREYTR QNPMATEKRLSLLEPYVPYMIHLADPDFTRSQDVQDLHDECLMFLMEVLTMRK ENNSHAFMKKMAENIKLIRDAQSPDESKTNEKLYTVCDVALCVINSALCNADSPKIK PVLPMKFFTOPEKVISLT"				
CDS	BASE COUNT	611 a	429 c	432 g	504 t
ORIGIN	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 7				

Db 445 GAAGGAGCGCCATGATGAGAGAGAGGCTGTCCAGTCCAGATCCGCTCAGCA 504
 QY LeuGluleuLeuValleuSerPheThrHisProIleSerPheHisSerAlaGluThr 669
 Db 505 CTGGAGCTTTTAAAGGTTGCTTTTACACATCTCACTGCTGCTGCGAGACACA 564
 QY PheGluSerLeuLeuAlaCysLeuLysMetAspAspGluValAlaGluAlaLeu 689
 Db 565 TATGAGTCCTGTTACAGTGCCTAAGAAAGAGATGACAAAGTAGACAGAACTGCAATA 624
 QY GlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAla 709
 Db 625 CAATATTTTGAAGAACACAGCCACAAAATAGAAACCTCCCTCCAGATACGGTCAACC 684
 QY LeuLeuProValleuHisIleLysSerLysGluGlyProProArgGlnAlaLysTyrAla 729
 Db 685 TTGATCCCATTTTACTACGAAAGCCAGCGGGGACCTCCACACCAAGCAAGACAGCT 744
 QY IleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGlu 749
 Db 745 GTTCAGTCGATCCATGCGCATCTCTCAACAGAGAGCTCCAGTCGCGACAGATTTTTCAG 804
 QY ProLeuHisLysSerLeuAspProSerAsnLeuGlnHisIleuIleThrProLeuValThr 769
 Db 805 CCACCTGAGTGGAGTCTGAATGCTGATGACACAGCACTTATTAATCCATTAAGTTTCA 864
 QY IleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaIleProTyrPlySerSerTyrVal 789
 Db 865 CTGGGCCACATTTCCATGTTAGCAGCCAGATCAGTTTCCCTCCGATGAATCTGTAGTG 924
 QY AlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThr 809
 Db 925 GCAAACTTTATTTGTTAAAGACCTTCTAATGAACGACAGCTCAACAGCTGAGAAAGATGGA 984
 QY LysLeuTyrValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIle 829
 Db 985 AAATATAGGCTCCAGATGAGAAAGTTTCTCCGAACTTGTAGCAAGGATACAGGCATTT 1044
 QY LysMetMetValArgTyrPheLeuGluMetLysAsnAsnHisSerLysSerGlyThrSer 849
 Db 1045 AAATCTGCTGTAAGTGGCTGTTGGGTATGAAAACACACAGCTTAATCTGCCCACTCA 1104
 QY ThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLys 869
 Db 1105 ACCTCTCGTATTATATACGCCATGCTGGTTAGTGAGGCGCACCTGACAGCAAAAGAG 1164
 QY IleSerLysProAspMetSerArgLeuArgLeuAlaIleAlaGlySerAlaIleValLysLeu 889
 Db 1165 ATCAGTAAATCTGATATGCTCCGTCGAGTATGAGCTGCTGTAGTCCATATGAAAGCTT 1224
 QY AlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnIleThrGlnLeuCysAlaLeu 909
 Db 1225 GCTCAGGAACCTTGCATACATGAAATATTACCCAGAACAGATTTCAGCTGTGCACTG 1284
 QY AlaIleAsnAspGlyCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGly 929
 Db 1285 GTTATTAAATGATAGGCTACCAAGTAAGACAGATATTTCGCCAAGAGCTTCAAGCA 1344
 QY LeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 949
 Db 1345 CTGTGTAAGTGTCTTCCACTGCGAGTATATGCGCATCTTGTGTTGTGTCNAAGAC 1404
 QY ProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArg 969
 Db 1405 CCTGTGAAGGAAGAGAGACACAGCGTCGCGAGTGTGTTAAAGAACATCAGATCCGC 1464
 QY ArgGlnTyrLeuLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuProGlu 989
 Db 1465 AGGAGATACATCAAAACAGAACCCCAAGCCAGAGAAATATTGTACAGCTGCCCTGAA 1524
 QY TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGln 1009
 Db 1525 TATGTGTTCCATACATGATTCACCTCTAGCCCATGATCTGTGATTTCCACACAGCA 1584

QY 1010 AspIleGluGlnLeuLysAspValLysGluCysLeuThrPheValleuGluIleLeuMet 1029
 Db 1585 GATCTTATCACTTCTGATATTAAGAGAGTGCCTGTGGTTTATGCTTGAAGCTTATATG 1644
 QY 1030 AlalysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGln 1049
 Db 1645 ACAAGAATGAAACACACAGCATTCATGAAAGAAAGATGGCAGAGAAATATCAACCTA 1704
 QY 1050 ThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCys 1069
 Db 1705 ACCAGACAGCCGCCAGTCTCCGATGATCCAGAACAAATGAAAACTTATACGTTTGT 1764
 QY 1070 AspValIleAlaMetIleIleMetSerLysSerThrThrLysSerLeuGluSerProLys 1089
 Db 1765 GACCTGCTCTGTGTGTATTAATAGTAAAGTCTTTGTGCAATGCGACACTCACCAAG 1824
 QY 1090 AspProValleuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThr 1108
 Db 1825 GACCCAGTCTCCCAATGAAGTCTTTACACAGCTGAAAGATTTTCCCTCACA 1881
 RESULT 13
 AL844559 216608 bp DNA linear VRT 28-JAN-2003
 LOCUS zebrafish DNA sequence from clone DKEX-14K1, complete sequence.
 DEFINITION AL844559
 ACCESSION AL844559
 VERSION AL844559.14 GI:28144648
 KEYWORDS HMG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 216608)
 AUTHORS Babbage, A.
 TITLES Direct Submission
 JOURNAL Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
 On Jan 30, 2003 this sequence version replaced gi:27801645.
 COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/C.elegans/wormpep>
 beginning 'Dr' were identified by the Recon repeat discovery system
 (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir'
 were identified by Rick Waterman (Stephen Johnson lab, WashU). For
 further information see http://www/projects/D_rerio/fishmask.shtml
 DKEX-14K1 is from a zebrafish BMC library
 VECTOR: pIndigoBAC-5.
 FEATURES
 Location/Qualifiers
 I. .216608


```
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-14K1"
/clone_lib="Daniokey"

BASE COUNT  68345 a 39077 c 40063 g 69123 t
ORIGIN

Alignment Scores:
Pred. No.:      6,79e-86      Length:      216608
Score:         1836.00      Matches:      916
Percent Similarity: 18.40%      Conservative: 148
Best Local Similarity: 15.84%      Mismatches:  215
Query Match:    25.52%      Indels:      4511
DB:              5          Gaps:         39

US-09-512-581b-2 (1-1391) x AL844559 (1-216608)

OY      37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyr 56
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      95986 TTGGGGTTAAGACATATATGACATGATGATCAAGACTCAGAAAGAGAAAGACAGATAC 96045

OY      57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspVal 76
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96046 TTGGCTTTGGCTCTTACCCTG6CCCTGAGTCTTTCTGCGAAACCCCAACAGGATGA 96105

OY      77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96106 CGGTACTGCTGGCTGGCTGCTGCTGCTGACATCTTCAGATCATATGCTCCAGAGGACCA 96185

OY      97 TyrThrSerProAspLysLeuLys----- 104
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96166 TACACCTCATGATAAACTCAAGGATTAACCAATACTCATCATCTAGCTGCTTGAGA 96225

OY      104 ----- 104
Db      96226 ATGTAAATCTTTCAAAACACTTTTACGTTGTCMAACTGATTTCTTCTACCTGCAC 96285

OY      104 ----- 104
Db      96286 ACAATCACCTCATCATACGAAACTAAATTTGATATATTTTCAGAGTTTAAACAGATAC 96345

OY      105 -----AspIlePheMetPheIleThrArgGln 113
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      96346 AGTGATGATGGTATTTGTTTTCCTATCTAGAGAAATTTTCTTTTCATCAGAGGACG 96405

OY      114 LeuLysGluLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuGlu 133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96406 TTGAAGGCTCTTGAGGACACCAAGCCACAGTTCAACAGATATTTTACTTGTAGAG 96465

OY      133 ----- 133
Db      96466 GTGAGATGGCTCTTTAATGAGAGATTATATGATCTTTTGTCTCTGTTTACAGAAAGAA 96525

OY      133 ----- 133
Db      96526 AGGCTTGATGCTGTTTAAATTACTCTCAGACCTCAGAGCATCTAGACGCTTTTTCG 96585

OY      133 ----- 133
Db      96586 ACCAGTTTGCTTTGGTTTATGATCGTATTATGATTTGTCATGACAGCTAACTTTTCAT 96645

OY      134 -----AsnIleAlaTrpValLysSerTyrAsnLecysPheGluLeuGluAs 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96646 TTGACTTGACACAGAACTTGCGATGGGTGAAGTCTCTACACATCTGTTTTGAATTGAGGA 96705

OY      149 pSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn----- 166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96706 CTGTAATGAATCTTTATCCAGCTCTCAAAACCTCTTTCTCTGCATCAA-GTGAGTTTA 96764

OY      166 ----- 166
Db      96765 ATTTGAGACGATGTAAATTATCCCAATTTGTTCTCTCAACTGTCTCTTCTGATCAGAC 96824
```

```
OY      167 -----AsnGlyHisAsnGluLysValHisMetHisMetValAspLeuMet 181
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96825 AATGCACCTTTTCACTTAACAGGCATATACAGAGGTGCAATGCAATGCTTTGGACTGAT 96884

OY      181 tSerSerIleIleCysGluGluAspThrValSerGlnGluLeuAspThrValLeuVal 201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96885 GAGTTCATCATCATGAGGAGGTGATGAGTGCACAGAGAGCTGTGACACAAATCCGCAT 96944

OY      201 lAsnLeuValProAlaHisLys----- 208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      96945 CAACCTCATCTCCCTGCTCACAAAGTAGGCAAAATGACCTATTGTGATTTATTTTACATTCCA 97004

OY      209 -----AsnLeuAsnLys 212
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97005 GTCTCAGATAGCTAACCAATATGCGTTTCACACCATTCCTGCTGCTTTAGAACTTGATTA 97064

OY      212 sGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyr 232
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97065 GCAGGCATATGACCTGGCCAGGACTCTTCTTAAGAGAAAGGTACAGACTATCGAAACCTG 97124

OY      232 rIleThr----- 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97125 CATCGCTCAGTGAAGTCACAAATATCTGTGTAATTATTTATCTTTGAGATAAGTTT 97184

OY      235 ----- 235
Db      97185 CAGCATTTAGCATGTCATATATTATTCATATGTTGATTTGCTTGCGATTTCTTTTCT 97244

OY      236 ----PhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHis 254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97245 GTAGTCTTCATATCAGGTGTGCTGATGGAAGAGCCCTGGGAGTGTATCTGACAGCA 97304

OY      254 sValPheAspLeuIleLeuGluLeuTyrAsnLysSerHisLeuLeuSerValLe 274
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97305 TGTATTTACCTCATATCAGGAACCTCTTGCAATGACCTCTGCTCAGTCTGTGAT 97364

OY      274 uProGlnLeuGluPheLysLeuLys----- 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97365 GCCACAGCTGAGTTTAATTGAAGTGACCAATTTGACTCAGACTTAATTTACACCAG 97424

OY      283 -----SerAs 284
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97425 CTTCATACAAAGGCTTTAGAGTAAGTCTTTGTTTGTGTTTGTCTTTGTAAGTAA 97484

OY      284 nAspAsnGluGluIleGluGlnValLysLeuAlaLysMetPheGlyAlaLysAs 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97485 TGATGGAGAGGAAAGTTAGCTGTGTAATAATGCTAGCCAAACTCTTGGCGCAAAAGA 97544

OY      304 pSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrIleGly----- 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97545 CTCTGAACCTGCCACTCAGAAATGCCCACTGCGAGTCTTCTGGGAAGGTAGGATGG 97604

OY      320 ----- 320
Db      97605 CTTTACATTTAAATAATCTGAATTAAGTGTAGAGATTGTGAGATTCTTAATGAA 97664

OY      321 -----ArgPheAsnAspIleHisValProIleArg 330
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97665 TTTGCTTTAATTTTTCCTTTTGTCTTGAAGGTCAATGACATCAATGATCTGCTGTTAG 97724

OY      330 gLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97725 ACTGGAATGTGTGAAGTTTGCACAGTCACTGCTGATGAACCATCAGACCTTGCACAGA 97784

OY      350 pLeu----- 351
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97785 TCTTAACAGTCTGTTCTCTGTAATACATGCTGTTGTTTCTATTTTAAGCACT 97844

OY      352 -----ThrGluTyrLe 355
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      97845 CTTGACATAGAGATGCTGTGTTAATTAACTTAATGATGTGCTATATTTCACATCAGAGTTTT 97904
```


Dd 100062 TAGTGTATACACAGACTCTAAGTAGAAGTATCCAGTAAGGCTTAAATCTGTAGA 100121
Oy 580 ----- 580
Dd 100122 ATACAGTGTTTAAACGTGTGATTCATCTATPAAAAAGTCGAATCATAGTCTTCAACG 100181
Oy 580 ----- 580
Dd 100182 AGTGGCTTGTATACGACTCATTTAGGTGTTTCCCATGACGTAGCAAGCAAAAGTTT 100241
Oy 580 ----- 580
Dd 100242 TTACGTGCACGGCGCAAAACCGGAGATTTCAAACCTGAGGCCCGCTGTGACGAGT 100301
Oy 580 ----- 580
Dd 100302 AACCCAGACACACACACACACACCACAAACACAGCAGACCCCAAAACACAGCACA 100361
Oy 580 ----- 580
Dd 100362 CAACATGCCGGTGCATTGAAGTCACACTGCAGATGATATATTTGAGTCTCTATCCAAG 100421
Oy 580 ----- 580
Dd 100422 ATGAACCTCAGATTATTAACCAAGCAGTTGAAACTTCTGGAAGCTACATGCTCAAG 100481
Oy 580 ----- 580
Dd 100482 AATACTTCATCTGAGTTTGTAAAGAGATCAGTAAGAGTACTGATGACGCTCAGG 100541
Oy 580 ----- 580
Dd 100542 ATGGGTTTCTTCTACATTCTCAAGTGAAGTACGTGCGGTTAAAGTTGTTCCTCGTT 100601
Oy 580 ----- 580
Dd 100602 TACTCTAGCTTCAAAATGATTTAGTTGATTTGTGTACTGTATACCGCTACTGTAT 100661
Oy 580 ----- 580
Dd 100662 CAGTTTAACTGATTTTGAATCTATGCGGTGCAAAAGTCACGTTAAAAACGCGACGCTG 100721
Oy 580 ----- 580
Dd 100722 CTGTTTGTATGAGCTCCGCTGCTAGAGTCTGCTCCCGGATTTATTCGAAATGTAT 100781
Oy 580 ----- 580
Dd 100782 TCCCGCGCCGAGAAATCTGCGCGGGAGAGAGAGAGAGAGAGCGGTAGGAG 100841
Oy 580 ----- 580
Dd 100842 AGCCAGCAACGAACGAGCTTTGTGTACTTGTGCTGTGTGTTTTTATACTGACAG 100901
Oy 580 ----- 580
Dd 100902 CTTCGCTGTCTGACTGTATACTGGGTGATTTGTTGTCTGTCGCTAGCGGAGAC 100961
Oy 580 ----- 580
Dd 100962 CGGGCGGGCGGCAAAACGAGCGGGTCTCAGCTAAACCTGGCGGGTGGGGCT 101021
Oy 580 ----- 580
Dd 101022 GAACGGGAGCGTTGTATCCGAGGTGTGTGTGTTTTTGTGTGTGTGTGCGG 101081
Oy 580 ----- 580
Dd 101082 AGTACAGACATGTCCTTCTGTTGGTTATCCGTTGCGACTATCCACTGCCATAGTGCG 101141
Oy 580 ----- 580
Dd 101142 CAGTAAGTCCAGCGCTACACTATGAGCGTGTATGAACGTGTACTTTTAAATTG 101201

Oy 580 ----- 580
Dd 101202 CTGATTAGCTATTTGGCCATTTCCCTCTCTGACTGAAGGAGCTCGACCAATGCGCAGAC 101261
Oy 580 ----- 580
Dd 101262 TGTCAATGCTCAATCAGCGGAGATTACCTTCGGCCTAAGAGGGGTTTGGGAACMAATG 101321
Oy 580 ----- 580
Dd 101322 AATCAGTGACGATTCATACAGAGTCCGTGGGATTAATAGATAAAATAATGACAGATT 101381
Oy 580 ----- 580
Dd 101382 ATAGACCATGAAGTGTGTTTTTGGACCTTGATGATATTAAGACTGTGTTGGAGACCT 101441
Oy 580 ----- 580
Dd 101442 TACAACGAAGATATGACCTTATTCATGTATATATATGSGGCTTTAACAAACCTTTTTT 101501
Oy 580 ----- 580
Dd 101502 AATGCTTGTCAAAACCTCGGCCATGCTTGTCAATPACCAGGGCTTCACATGAACTTTTTT 101561
Oy 580 ----- 580
Dd 101562 GATCACCAGCAACAGTGGCTATTAGTTTTTCAACGTTAGTACCATCTGCAATTTACT 101621
Oy 580 ----- 580
Dd 101622 AGCCACATTTTGTGTGGAAAAATATTTTATATGACAAATTTACTTTGACATGC 101681
Oy 580 ----- 580
Dd 101682 TAAATPACTGATTAGATTTTGTGTTTGTTCACATGCGCCCTAATTAATGACGCTT 101741
Oy 580 ----- 580
Dd 101742 TTGTGTGCTGTATGACAAATGATCATGCTTCAATGTTCTATTAACAATTAGTTT 101801
Oy 580 ----- 580
Dd 101802 TTATTTACAGGATTTTTCAGAACTCAAATTTAGGAGTACCAATTTATCTGACCA 101861
Oy 580 ----- 580
Dd 101862 CTCGACACATTAATTTATTTAGCCACACGTTTTTAAATGCGTCMAAACAGCTAATAT 101921
Oy 580 ----- 580
Dd 101922 AGCTGTATTAATCTTTTATTTAACAAACATTTCTTAAAAAAGAAAAAAG 101981
Oy 580 ----- 580
Dd 101982 AAAAACATGACTTTTAAAAATAGTATATGATCATGATCAAAAATATGAGAGTAATCT 102041
Oy 580 ----- 580
Dd 102042 GTCCCTCGTAAAGTCTCTCAGATCAGTACTACTACAAAGTGAGAGTTAATCTCATA 102101
Oy 580 ----- 580
Dd 102102 TTTAAAGCAATGTGTTGTAATAAAATGATGATAACATATTAACAAAATAACCGTAA 102161
Oy 580 ----- 580
Dd 102162 CAAAGAAAGCAGGTGAATAAATATACCTGTGATATGAAGAGATGATCGATGTTTG 102221
Oy 580 ----- 580
Dd 102222 GCCAATTAATATCTGTTCAAAAGTGAAAGCAACTGTGCTGTACAAAAAGACCAATCT 102281

QY 580 ----- 580
Db 102282 GGAGCTCTGAAAGCAGCAGACTGCGGTGCTATTACCATTAATTTTGTCTGTCTATTA 102341
QY 580 ----- 580
Db 102342 TGAAGAGAACTGATTAACAGTTCGGTCCAGGCAGATTGCTTCAAGCAAAACAAGAGA 102401
QY 580 ----- 580
Db 102402 GCGGACACGGGTTTATAAGTGGCGGTATCATCAGCTGTGTGCAGAGCTGATCATCT 102461
QY 580 ----- 580
Db 102462 CTCATTCATTTCAGTCCAGATCATCCAGTATTACCTGCTTCTATTGTGCGCGGAAA 102521
QY 580 ----- 580
Db 102522 GCAAAATTATTTTGTAGTGTGCTGCTTTCTCGATTCTTAAGATCACATTTGCACGCATAT 102581
QY 580 ----- 580
Db 102582 TGGGTACCTTATAGTCGAACCTGCTTTTAAGAAAACATATTTAAAAATTTGTTTC 102641
QY 580 ----- 580
Db 102642 AACCACTGAAGTGGCTAGTGGGTAGGCTGCACGATATATCGTTTCAGCATCGAAATCG 102701
QY 580 ----- 580
Db 102702 CGATGTCCGATCCGCGATAGTACATCGCAGAGGGTGGCATGCCAAAGCAAAAAAAA 102761
QY 580 ----- 580
Db 102762 AAAAATTCATATATATTATGAGGTGAGCGCAACCGAATATGTATTTCAGAAAAGCAGC 102821
QY 580 ----- 580
Db 102822 AATATATGTTTTTACGAATCTATTTCAACATACCTGATTAATTTTGTATTTGGG 102881
QY 580 ----- 580
Db 102882 AACAGAAAAACATTATATCAAAAAAGCTGCGTTTCTTAATGAGACCGCAGTCATGCCG 102941
QY 580 ----- 580
Db 102942 CGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGTGGGTGGGGCCGGGTA 103001
QY 580 ----- 580
Db 103002 AGGATCTGAGCAAGCTGCTTTCTATACGAAAGAGAAAGGCTTGGCAGAGCAAAAA 103061
QY 580 ----- 580
Db 103062 TACTTTTCTGCATTACTATTTTGTATTATTTATCTAATAAACAAGTGAAGAG 103121
QY 580 ----- 580
Db 103122 TTCACTCAAAAGTAAAGTCTTCTGTATTAATTAACAACCTCCGCTCATTCCTACT 103181
QY 580 ----- 580
Db 103182 ACTTTCAGAACATTAATTTGATATTTTACCATTGAATAATGCAATGCAATTAAGAAAT 103241
QY 580 ----- 580
Db 103242 TGAATAATTAAGCAATACACAGATCTGTCCCGCTCAGGTACGGGTGCACCTGTAAAC 103301
QY 580 ----- 580
Db 103302 CTTAAAAAGCAGACGGGGGAGGGCGCACTAAACTCAACAAGGAAAGAACTGGGCATTC 103361
QY 580 ----- 580

Db 103362 AATTCATGTTTGTGTTAAATTAATCTTAATAACAGCTCTTTGATCTTAAGGGAAAA 103421
QY 580 ----- 580
Db 103422 AAAACGAAAGAAATTTGTGCTTAATTAATTTAGCAAAAGCAGCACTGCTGCATTT 103481
QY 580 ----- 580
Db 103482 TTCATTTGTTTTTCAGTTAGGGATTTGGCTATTAGAGTAACCTAATAAATTTGTTTT 103541
QY 580 ----- 580
Db 103542 GTTGGACACGTAGTCTACACAATTAATTAATTAATTAATTAATTAATTAATTAATTA 103601
QY 580 ----- 580
Db 103602 AAACTAACCTATACATTTGATTAACAAAAACCCCTTTTAATGTTTGTCTTCATTTTGACA 103661
QY 580 ----- 580
Db 103662 CCGCTACAGTTGTAATAGAAACATCATTAATGCAAGACATGCTCAAGTTAAA 103721
QY 580 ----- 580
Db 103722 AACACCTTTGAGTTTAAAGATGATCAGCTCGAAAGTCAAGTTACAGTGTATTAATG 103781
QY 580 ----- 580
Db 103782 AAATGGTACATGATGAAGGGGATTTGGCCAGCACACAACCTCATGCGCTGACGATA 103841
QY 580 ----- 580
Db 103842 GCTGAATGATTAATCTTAATAAATTAATAGCTATGCAATACAAACTTTAAACATTTAT 103901
QY 580 ----- 580
Db 103902 TATTTTTTAATGAAGAAAAAGAAACAGAGAGTAATAATTAATTAACACACTGCTCT 103961
QY 580 ----- 580
Db 103962 GGGCGTATCTTATTTGTTTGTATCCAGAAGAGGGGGAAGAAAGCAGATGAACCTGG 104021
QY 580 ----- 580
Db 104022 CTGGGAGAGGGGATTAACATATCATCTTTCAGCAAAAGATCAGACATCTGTGATAC 104081
QY 580 ----- 580
Db 104082 GTTTATCATCAACGACAGTTCGTGTTTAATCATTAATTAATTAATTAATTAATTCCT 104141
QY 580 ----- 580
Db 104142 CTCTAGACTTAGTTTTAGGGCTGTCTTGTGGCGCAATATAGCTCGGCTTATTTGCAC 104201
QY 580 ----- 580
Db 104202 GTACGAATCCGATGTGAAGAATTAATAAGCGTGCATTTTATTCATCTTTTCGTA 104261
QY 580 ----- 580
Db 104262 AGTTATTCATTTTTTTTTTAAGTGTGTAACATTTTTTACTGGCAACATGTTCCATGAC 104321
QY 580 ----- 580
Db 104322 CTCGTACATCAACAATAATAGTACCTACTTAAGTTAATTTACCTTTTATGTAAAAATTA 104381
QY 580 ----- 580
Db 104382 AGGAATACTTAATTAATAATTAATAATAAACAATCCGATTAAGATGATGTTATTTT 104441
QY 580 ----- 580

[illegible]

QY	580	-----	-----	580
Db	105582	ATAATATAATCGCAATATATATCGCAGGGGAAAAAATATCGAATGTCAATTTTTTCC	105641	
QY	580	-----	-----	580
Db	105642	AATATCGTCGACCCCTACTAGTGGAGTGAAGTCTTAACCTGCGCAAGCTAAATCTTACC	105701	
QY	580	-----	-----	580
Db	105702	CACATTTGGCGGTTGGCGGTTAATGTAAAGCCCTGTCATTAACCATTCACAAATCGT	105761	
QY	580	-----	-----	580
Db	105762	CCTGTTAAAAATGATTACAAATGACCTATGTCTTGATCAACAATAAAGCAAGCTTTATTA	105821	
QY	581	-----	-----	581
Db	105822	TAGTACTAAATTTAAACAGACATTAAGTATTGTAAGTGGCAGTGTCTTCACAGGGAA	105881	
QY	583	IllethrlslyslsleuglyasnProlyslnProthrasnProphelenglumetllys	602	
Db	105882	ATCACACCAAAAGCTGACATTTCCCTTAAGCGCGCAAAATCCATTCCTTGAGATGTGTAA	105941	
QY	603	PhelenuengluaqylielalProvalhistleaspthrglusertile	618	
Db	105942	TTCCCTGCTGGAGAGAAATTCGACCGATGTCACATTGACTGTGAGGCAATTAAGTAATGAGA	106001	
QY	618	-----	-----	618
Db	106002	ACTGTTTTTTTTTCTGTGGAAAAAGTTTCATATAGCAGTCATGCCGACATGAT	106061	
QY	618	-----	-----	618
Db	106062	AAAGTATGTCTTAACTTCTTTTTTATTATTAGTTTAATAAGATTGGCAGCAACAAAAA	106121	
QY	618	-----	-----	618
Db	106122	AAAAAGATTGAAAAATTTTATTTCATATTAATGAAGTAATATCTGATGTTTAAATCT	106181	
QY	618	-----	-----	618
Db	106182	TTCAAGAACTAATATAAAAAAGCTTTCGTATTAACAGCAATCAATTAATGACTGTACATG	106241	
QY	618	-----	-----	618
Db	106242	GACAAGATTAAAGTAACTGACACACACATTTGCTCATGACCTCACTTTCAGTTATACAA	106301	
QY	618	-----	-----	618
Db	106302	AACAACATGCAAGTACCTTATGTGTGAACATAAATTTGTTCACTTACAAAAATAAGTGT	106361	
QY	618	-----	-----	618
Db	106362	TCAGTCACTTTTAATCAAAATTTGTAATATATACAGAGTGTCTGTTAAATGCACTTCTAT	106421	
QY	618	-----	-----	618
Db	106422	GGAGTCTATTCAGTCACCTCGACAGAGTGAGAGTGAACATGAATGAGAGAGAAAGAA	106481	
QY	618	-----	-----	618
Db	106482	ACTGACTGAGTGGGAGAAAAGATAAAGCACTATATGCGCCATGTCTTTCCCTCCCTAAC	106541	
QY	618	-----	-----	618
Db	106542	GTTTGTATTAACAGCCTCAATGGCTTTTAAAGATGTGACATCTAGTGAAGAAAAGTGA	106601	
QY	619	-----	-----	619
Db	106602	ACTGAGGTATTTGCTGGCCTTTTTTGATTAACACATTAATGCTGTCTACACACTGCGCTGT	106661	

[illegible]

Db	1077142	GCATATTCGATTCGACGAGATGATCATCTGAATGATCTGAATAATGCTTTGAACTAGACACTA	1078001
QY	755	-----LeuaspProSerAsn-Le	760
Db	1078002	AAACCCAAACTTGTGAAAGAATCATATCACTACAGGTAAATATGATCCAGCACAATTAT	1078601
QY	760	u-----	760
Db	1078602	CACCTCTAGAAAGATCTTATTTTATTTTATTTAGATCATCTTTTTCATCATGATGTGTTTC	1079211
QY	760	-----	760
Db	1079222	TACAACAACAGCTTTGTTTCAGGCGATGAAGAATTAAGTAATGAGAGTGGCCAGCTCAATC	1079811
QY	760	-----	760
Db	1079802	GCCTTACTTCAACCTAATTTGAACACTCATTTGGCTAAATTTAGGATTTGGCGATAGACGAT	1080411
QY	760	-----	760
Db	1080402	TGTATCAACGATAGGCAAAAATATGCGACAGAGCTGAACCTATGATGATTTTAAAGACG	1081011
QY	760	-----	760
Db	1081002	ATATTAGCTCCTTTCGACGCTGTACATGTTTATTTAATGCAATTATAAATAAATGAATG	1081611
QY	760	-----	760
Db	1081602	TATTACTATTTTATATCATCATATTAATAAATAAATAAATAAATAAATAATGATTTGCTT	1082211
QY	760	-----	760
Db	1082202	CAGCTGTTTATATCAACAATCACCTGACAGACACTTTTGGCAAAAATGATTTTTCAT	1082811
QY	760	-----	760
Db	1082802	TGCTCTCTATCAAAAAAAGCTTTTACTGCCATCGAAAAAGTTTATAGTCAAAATAGGCTAT	1083411
QY	760	-----	760
Db	1083402	TACTGACTCGCCGACGCCATGTCTAAATGAACATATCATCATATATGGCGATCTGAAAA	1084011
QY	760	-----	760
Db	1084002	TTTTGTATATTTGCCCAACACTAGCTGAATACCTGTTTCTAGGTGCCAGAGTTATATTA	1084611
QY	760	-----	760
Db	1084602	AAAGCAATATTAATAATATTCAGATAAATATTAATATCAATATTTTTCCTPAACTGAA	1085211
QY	760	-----	760
Db	1085202	ACACTAATATTAATAATTAATTAATGACACCTTTTCTCTATCTTGTGTGATGACACACAGT	1085811
QY	760	-----	760
Db	1085802	GTCCTGCTTCTATATAATGAATAAATAGTCGACTGACTTTTACTGCTGATTTTTCG	1086411
QY	760	-----	760
Db	1086402	ACAGTAGATTTGCTCCCTGTACACCTGATGSGTGAAGAAGCACTGACTGATGCTGACATCT	1087011
QY	760	-----	760
Db	1087002	ATTATGATCAATCTAATAATCTTCATAGAGTGGCAGACGTCTCTATATCAGACACAAA	1087611
QY	760	-----	760
Db	1087602	AAGCCACATGTGTGAAAGTTTTTACAAAAATGAAAAATGGGTTGTACTACATTTGTTTGTG	1088211
QY	760	-----	760

Db 108822 CTCGCAATTTCTACCAATGTATTTATGATGCAAGTTTACATTTCTGCTACATAT 108881
QY 760 ----- 760
Db 108882 GAGACTAAATGGTTGTAATTTACAACCCCTGACTATTTTGTATCTTGTGTAGATAG 108941
QY 760 ----- 760
Db 108942 AACATGACACTTTATTTATGAACATCATGATCATCAATTTAAAGTGTCTTTGGCA 109001
QY 760 ----- 760
Db 109002 CATCATATCTGAGGTGTGACGAGATTTTAAACTAGATATTGCAAAAAATAAATA 109061
QY 760 ----- 760
Db 109062 AAAACAAAAGCAATTAACCCACTTTCTCTGCAAAATATCAGCAGATTATATATAAATG 109121
QY 760 ----- 760
Db 109122 GGTGCTAACTTTGCCCTTTTATGATGAATAATGTGACAAATGTATTTAAATGTTTAA 109181
QY 760 ----- 760
Db 109182 TTTATTTTATTTTATTTTATGCAATTTGTTTCTGACAGATTTTCAAAACATTTTAA 109241
QY 761 ----- GluHisLe 763
Db 109242 TCTCTGTGTTTGTCTCCAGCCCTGTCTCGAGTGTGAATGCAAGATGTAACCTGACGT 109301
QY 763 uLeuThrProLeuValIhrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAl 783
Db 109302 GATCATCTCAGTCGTCTCTCTGAGACACATCTCCATGCTGGCCCGGAGATCATTTGTCATC 109361
QY 783 aProTrrpysSerTrpValAlaThrPheIleValIlysAspLeuMetAsnAspArg-- 802
Db 109362 TCCCATGAAGTCAAATTGTAGCTAAATTTATGTAAAGACCTTGTGATGAATGACAGAGT 109421
QY 802 ----- 802
Db 109422 GAGCATATGTATGTATTTATCTGCTTTCAGCATGTGGAATGTTTTTATACCTTCAT 109481
QY 803 -----LeuProGlyLysLysThrIleLysLeuTrpVal 813
Db 109482 GTTAATTTTACCCCTCTCATCTCTTAGTCTGTG- GGAACAAATAAATGGCGCTCTGAGC 109540
QY 813 lProAspGlnGluValIserProGluThrMetValIys----- 825
Db 109541 AGCTGATGATGAGGTTTCTCCGAAAGTTTGGCTAAGTAAAGTAATATATCATATA 109600
QY 825 ----- 825
Db 109601 AAATTTCACTTTGTCCCTTTTAAATAATCCCAAGACACTTTCACCTGAGAAAATACAGTT 109660
QY 825 ----- 825
Db 109661 AATACATAATTAAGGTAGAAAACGTCAATTTTGTATGTCTATTTATTTATTTGTTGTCT 109720
QY 826 -----IleGlnAlaIleLysMetMetValArgTrpLeuGlnLysMetLysAsnAsn 843
Db 109721 GTTTCAGGTGACACCCATTAAGCTGTGTGTCCTGCTGTTGTAGGAATGAAGAATAATC 109780
QY 843 lAsSerLysSerGlyThrSerThrIleuArgLeuLeuThrIleLeuHisSerAspGly 863
Db 109781 AATCCAAATACAGCTACTCTCTCAGACAGCTGCTCTCAGCCATGTGCTCACTGAGAGAG 109840
QY 863 sPLeuThrGlnGlnLysIle----- 870
Db 109841 AACTCAGGAGCAGAGAAGAGTATAGTACTACTAGTATCTAGAGTATTTAGTATGAT 109900
QY 871 -----SerLysProAsp 874
Db 109901 AATAAAGACATATGCTCTAACACACATTTGTATGTTTGTGTACACAGCAAGTCTGAC 109960

QY 875 MetSerArgLeuArgLeuAlaIleGlySerAlaIleValIlysLeuAlaGlnGluProCys 894
Db 109961 ATGTCTCGACTAAGGCTTTGACCCGGAGATGCTATCTGAAATCTAGCACAAAGCCCTGT 110020
QY 895 TyrHisGluIleIleThrLeuGlnGlnIleuArgLeuLysAlaLeuAlaIleAsn----- 912
Db 110021 TACCATGACATATTACTCTCAGAACAGATTCCAGCTCTGTGGACTGTCTATTAATGTGGT 110080
QY 912 ----- 912
Db 110081 TCAGCTACTGGGCAATATATCACTCAATATTTAACTGAATAACAGAGATGACCTTGAC 110140
QY 912 ----- 912
Db 110141 TTTTAGGACGACAAAAAATAATTTCCCTTTGCCACATGATGACGTATATTCAGGGTC 110200
QY 912 ----- 912
Db 110201 TGTGCACATCAATATAGCAGGACATTTTGTTCAGATGTTTGTATGATGAATGGCTGAA 110260
QY 912 ----- 912
Db 110261 GATCAATTTTAAAGTAGACTTTTCTTTTCTTTTGACAGCTCTTTAGATTTAGAC 110320
QY 912 ----- 912
Db 110321 TAGACCTGTGTTACACACACACACAAAGCAATTAATAAATAAACTCTTAACTAAA 110380
QY 912 ----- 912
Db 110381 TTGTAATAAAGAAATCGCTGTATATAGCTTATTTGACCTGTAGATTAATGGCTAAA 110440
QY 912 ----- 912
Db 110441 TGTGTCTCTGTGACTACATGTCTACAGCAACACTCTTCATTTATAGCTTTCTCT 110500
QY 913 -----AspGlnCysTrpGlnValArgGlnValPheAlaGlnLysLeu 926
Db 110501 CTTTTCCTCTCTTATATAGATGATGCTATATAGTGTGCGACAGATCTATGCTCAAAAGTT 110560
QY 926 uHisLysGlyLeuSerArgLeuArgLeuProLeuGlnTrpMetAlaIleCysAlaLeuCy 946
Db 110561 ACATGTGTGCTGTGAGAGATGTGTGTACCATGAGTACATGCGCTGATTTGCTTAG 110620
QY 946 sAlaLysAspProValIysGlnArgArgAlaHisAlaArgGlnCysLeuValLysAsnI 966
Db 110621 TGTCTAAGACCCAGTAAAGAGACGACGTCTCATGTCTGACAGTGTGCTGAAGACAT 110680
QY 966 eAsnValArgArgGluTrpLeuLysGlnHis-Ala----- 977
Db 110681 CAGTGTCCGACAGAGTACATCAACAGAACCTTATGGCTCATGTAGTACTTTTGT 110740
QY 977 ----- 977
Db 110741 TTAATTTTAAACATATATGTTTGTACTGTATACCATAAACACACCGGCAAT 110800
QY 978 -----AlaValSerG 981
Db 110801 TTCTCGGAGAGCTGTGAATGTAAAGATCTATCTAACTAATCTTTTGTGCAATTTGAG 110860
QY 981 lLysLeuLeuSerLeuLeuProGluTrpValValProtyrThrIleHisLeuLeuAla 1001
Db 110861 AAAAGCTCTGCTCTCTCTCCAGATATATGTGTGCCATACATGTATCTTCTAGCTC 110920
QY 1001 lAspProAspTrpValIysValGlnAspIleGlnIleuLysAspValLysGlu----- 1019
Db 110921 ATGATCCAGACTTACTTAACACACAGATTTGGACAGCTTAGAATGTCAAGA- GATC 110979
QY 1019 ----- 1019
Db 110980 GAGATCATATTTTGGCATTTTCTTTTCACTCAAAATGGCTGACTAATACTGCCAATCT 111039

QY 1019 ----- 1019
Db 111040 TTATACATTTTAAATATATATTTCTACTGAAACAGATAGATCTAGTTCCCTAATTTAT 111099
QY 1020 -----C 1020
Db 111100 GTTATGTGCTTTTAAATAGCAACTAATTGAATGAAATAATCTAATTTTTCACAGGT 111159
QY 1020 yslLeuTTPheValLeuGlu1lleuMetAlaLysAsnGluAsnSerHisAlaPheI 1040
Db 111160 GCTTATGCTTCATCTGAGTGTGTTGATGACAAAAAATGAAATATACATCCTGCTTCC 111219
QY 1040 leaArgLysMetValGluAsn1lleLysGlnThrLysAspAlaGlnGlyProAspAspAlaL 1060
Db 111220 TAAAGAAAGATGCTGAGAAACATCAACAAAGAAAGATGCTGAGTCCAGATGACCCAA 111279
QY 1060 ysmetAsnGlu----- 1063
Db 111280 AGGCAACGAGAGTTTGTCACTTACACGTTTATGACAGATTTGTATTTCTGAAATCC 111339
QY 1063 ----- 1063
Db 111340 GCTTGTAAAGTGTGACGTCAACCCGAGCGGTTCCAGTCCAAAGCGCTCAATTCACGTGA 111399
QY 1063 ----- 1063
Db 111400 ATGGGAGACTCATGAATGCTAATATAAACCTTTTGCAAAAGTATTAATCTTTCGA 111459
QY 1063 ----- 1063
Db 111460 AAATCTCGATCGCAGTATATATGTCCATGCCCTAATATCCAGTGGCCAAAGTGAATTAAT 111519
QY 1063 ----- 1063
Db 111520 TTTTATAAATVTTGTAATTTTGGTATTTGTTATGACAGAACCCAGAGATTTGTGTGTAC 111579
QY 1063 ----- 1063
Db 111580 ACTATGATTTTATATATAAATTAACCTTAATGTGTGATAGCAATAAACGTGATCATTAAC 111639
QY 1063 ----- 1063
Db 111640 GAATGATTTCTTCACTCAATGAATGCGGCTTGAGACCAGAAACATATTAATACATACGTC 111699
QY 1063 ----- 1063
Db 111700 ACAACACGTCACCACTTAACAAGCGGATAGGTTGTTAAACTTGTGATTAATTAATCTTTAC 111759
QY 1063 ----- 1063
Db 111760 TGAACGAAAAATGTGACTACATTTCCACTTGTCTTTTGGAGTACTGTGAGTGTATTTATG 111819
QY 1063 ----- 1063
Db 111820 AACACGTCCGATTAGCCATGTGTGCTCATGTGCTAATCTAGTCTAGTATTTGTA 111879
QY 1064 -----LysLeuTyrThrValCysAspValAlaMetAsn1lleMetSerLys 1080
Db 111880 TTTTATTTACAGAAACTATACATGATGTATGTATGTGATTTATTTGTCGCGCAACAAGA 111939
QY 1080 erThrThrLysSerLeuGluSerProLysProValLeuProAlaArgPhePheArg 1100
Db 111940 GCACTTATGTCACTGTGAGCTCTCCAAAGATCTGTGCTGCCAGCAAGTTCTACACC 111999
QY 1100 lnProAspLys----- 1103
Db 112000 CACCTGATTAAGTAAACAACACACTCAGACACCACTCTTACATGTGACCATTTCT 112059
QY 1104 -----AsnPheSerAsnThrLysAsnTyrLeuPro 1113
Db 112060 ATGAAGCTGCTTCTCTCTTTCTTCAACAGAGAGTTGTCAACAGTAAAGCAATATCTGTCA 112119
QY 1114 ProGluMetLysSerPhePheThrProGlyLys----- 1124

Db 112120 GCAGAACCTAAATAGATTCTGCAGACAGGAAAAATCCAGGTAGTTTATACCTATATA 112179
QY 1124 ----- 1124
Db 112180 TGCACATGCACACAGTAAACATTTGTGTAAATTAATCCGTATATAGCATGCGTAAAGC 112239
QY 1124 ----- 1124
Db 112240 ATGATTTTGTTTTACTAATATTTTCTGACATTTCAATTTAAAGTAACACATTTTAC 112299
QY 1124 ----- 1124
Db 112300 GTTGTAGACAGAGGTAGAAACCTTAGAGATTGGAGTTAATGAATAGGATATAGCA 112359
QY 1124 ----- 1124
Db 112360 AAAATTAATTTCTTACATTTGTTTGAACCCACACACTAACCAAGAAACGACATTAAT 112419
QY 1125 -----ProLysThrTh 1128
Db 112420 CTTTGAAGTAAATTAACGCAACCTTTGTGTCTATTTTCAGCAACCCAGCAAGCAGAC 112479
QY 1128 rAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLy 1148
Db 112480 AGGCTTCTGGGGCTGTAAACAACCTTTTAAACAGTAAACAGCTCGCGACCTGATCAATTA 112539
QY 1148 sSerSerArgMetGlnThrValSerAsnAlaSerSerSerAsnProSerSerProGly 1168
Db 112540 AACATACCTCTGAGACAGAGGAGCAATGCAAGCACCACACTCCACAGCTCACCTGC 112599
QY 1168 yArgLleLysGlyArgLeu----- 1174
Db 112600 AACAAATTAAGACAGGTTTGTGTGATCAGACATTTCTATTAAGATTAATGATTA 112659
QY 1175 -----AspSerSe 1177
Db 112660 TAATGACGATCTTCAATAAAGATTTTATTTCTTTCTATTTGAAGGATGTGAGTTTC 112719
QY 1177 rGluMetAspHisSerGluAsnGluAspLysPyrThrMetSerSerProLeuProGlyLysLy 1197
Db 112720 TGAAGTGGGGGCAAGGAGAACGA----- 112744
QY 1197 sSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgLys 1217
Db 112745 -----GAGATCTCTGTGATCAC 112761
QY 1217 glyLysLysThrProValThrGluGln---GluGluLysLeuGlyMetAspAspLeuThrLy 1236
Db 112762 AAAAGCAGTCAGTGTAAAGAGATGAATAAGTAAGTGTGTATGTTTCACACTGCTATC 112821
QY 1236 s-----LeuValGlnGluGln-----LysProLysGly 1245
Db 112822 ATCCTCAACTCAGCTTATTTGTTTGTCTGTGTAGAGAAAGCTGCTCAGCCAGCGG 112881
QY 1245 ySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnIntrPr 1265
Db 112882 G-----AGGAAGCGTGTGCTCTCTGCGCAGTAAAGGAGCAAGAA----- 112918
QY 1265 ogLglLysArgLeuLysGluAspLysPheLeuGluAsnGluAspLysGluGlnInserProp 1285
Db 112919 -----AACAGTGTGTCTTAAACCCCTCTGACAGTTTCACAGCCACC 112959
QY 1285 o---LysLysGlyLysArgLysArgProProLysProLeuGlyGlyLysThrProLysGly 1304
Db 112960 GCTCAATTAACCAACAGAGCGGTGACCCCTTAAACAGTGGCGGGGTGCCACACAGCA 113019
QY 1304 uGluProThrMetLysThrSerLysLysGlySer-----LysLysLysSerGlyPr 1321
Db 113020 AAAAGAGCTGTGAGCTCGACAGAGGAGCGGACGCGGAGAGAGACAGACACAC 113079
QY 1321 opProAlaProGluGluGlu---GluGluGluGluArgGlnInserGlyAsnThrGluGluLys 1341

Dh 22924 -ATTACAAACAAAATCTATGATATCATTTACGAGCTTAACCCGATCAATGACAGCTTGT 22981
Qy 270 euleuSerValIleuProGluLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgL 290
Dh 22982 TGCTGTACGTCTACTCTACAGTGGAGAACAACTGCTGTCCACGAGAGATCTGAGAGGC 23041
Qy 290 eu----- 290
Dh 23042 TAAAGTAGTAATTATGACATGTACCTTATTTGCAATTATTAACAATATTTTCGTTA 23101
Qy 291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAla 309
Dh 23102 CTACAGAGGCAACCACTCTCTGTCTGTATGTTCTCCGAAAGAGCTCCGCAACTACGCA 23161
Qy 309 ergLAsnLysProLeuTrpGluLncSTyLLeuGlyArgPheAsnAspIleHisValProI 329
Dh 23162 AAAAGTACCCCAACCTGTGTGAANAATCTTCTTGGGGCTTCTGGACATCATCCGACAG 23221
Qy 329 leatGLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAla 349
Dh 23222 TTCCATCAACAGTGTGTTCAGTCATCATGCACTTCCGTCMAATCACCACCAAGTCTCAGC 23281
Qy 349 ysaAspLeuThrGluTyLLeuLysValArgSerHisAspProGluGluAlaIleArgHis 369
Dh 23282 ACGTATCACCGAGAAATTTGGGCTTGGAAATCAGATCTTGACAGAACTGGTCCGACAG 23341
Qy 369 spValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp--- 387
Dh 23342 AGGTGTAATGGCATTTGTGGAACCTCCAAAGCGCGACTTCACTCGTTCTCGAAGCGC 23401
Qy 388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysG 407
Dh 23402 CCGATCTACAGAAATTTGTGGCGAGCGGAGCTAGATAGAACATACAAATTTGCGAGGG 23461
Qy 407 LuAlaMetMetGlyLeuAlaGlnIleTyLysLysTyLAlaLeuGlnSerAlaIleArgL 427
Dh 23462 ATGGATGAATGCTCTGGCTCTACATCTACAAGCGCGCAATTTGGGAACCCACAGATCTAA 23521
Qy 427 ysaAspAlaAlaLys--GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyLArg 446
Dh 23522 GCACTGGCTTAAGTACAGGGTGTACGTATTAAGAACAGATCATGATGATACATCA 23581
Qy 446 lAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyLMetValProH 466
Dh 23582 AAGTGGCTGTGGAGACCGCTGCTGTGGAACGCTTATCATCTGCTGCTGCTTCCCT 23641
Qy 466 lAsnLeuGluThrThrGluArgMetLysCysLeuTyLTrpLeuTyLAlaThrLeuAspL 486
Dh 23642 ATTAACTAGCTCCGGAAGAGCGCATGAAGAAGCTATATCATTTCTAGCGCATCTCGATG 23701
Qy 486 euAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
Dh 23702 CCAATGCCACCAAGCTTTGTGTCAGCTGAGAGAACCAATGAAGACAGCAATACG 23761
Qy 506 alLysAspLeuLeuAspLeuLysGln-----ProLysThrAspLaseArv 522
Dh 23762 TCAGTGTGTGATCAAGCTACACCACTCTAAGAGATTACGCGCGGTACTGTAGTACG 23821
Qy 522 alLysAla----- 524
Dh 23822 TCACGCGCAACGAGCCCAACATTGCAAGTGGGTTCTAATAAATTTCTTTGGATAGT 23881
Qy 525 --IlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGln 543
Dh 23882 TTGTAAACAGATGATGTTTTTTTCTTCTAGACTGCTCCAGATCCCTTAAGAGCGGCA 23941
Qy 544 AspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeu 563
Dh 23942 GAGTACCTAACCCAGTTAGCAACACACCTGCAAAAGATGCTCAGCTCTACGATGAT 24001
Qy 564 GluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgLLeu 583
Dh 24002 AACATGTCTTAAACGCGACGATAGCTGTGCGAGAGTGTGCTGACACAGATGGAGTTCTC 24061

Qy 584 ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPhe 603
Dh 24062 CTGAAAAGCTTGGCGGCCACAGTCCAA---TCGATCTGTATTTCAACACAGTTAAATG 24118
Qy 604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuLys 623
Dh 24119 CTGATGTAGCGCGGTGATGATGATGTGACAAAGAGTCAATTTGGCGCTATTATGAG 24178
Qy 624 -----Gln 624
Dh 24179 TAAATGCTTTTAAAGTCCCTTTTGCACACTTAATCCGTAATCTTTTGTCCAGCTTA 24238
Qy 625 ValAsnLysSerIleAspGlyThrAlaAspAspGluLysAspGluValProThrAspGln 644
Dh 24239 ATTGACAAATGCAATCGAAGAAAGGTCCATGTGTGAAGATGCGAATTTGGGCTCAGGAA 24298
Qy 645 AlAlaIleArgAlaGlyLeuGluLeu----- 653
Dh 24299 GCGGGCAGCGCGGACATCAAGTTGTGGCCGTGATAGAGATAATTAAATTAGTTATGATA 24358
Qy 654 -----LysValLeuSerPheThrHisProIleSer 663
Dh 24359 CAGAAATCCATTGATGCTTTATACCTTGACAGATG-CTTTCATACGCTTTTGGCGTCAT 24417
Qy 664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLys 683
Dh 24418 TTCTTCAACGATCTTGTGCTGCTGCTGATTAATTTCCCTGCTCAGATTCAGACAGATAT 24477
Qy 684 ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly-----SerLysIleGlu 700
Dh 24478 GTTCCACACCTGCTGTGAAGACACTAATCATATGATGGGCGCTATCAGCCACTGATGAC 24537
Qy 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLys 720
Dh 24538 GACCTTACACCGGCTATTTCTGAGAGATGCTGCTGGGTGTGCAAGACTTCGCACTTAT 24597
Qy 721 GlyProProArgGlnAlaLysTyLAlaIleHisCysIleHis----- 734
Dh 24598 GGAACCCCAAGCAAGCAAGACGCGCTGAGTATTTTCTGTAACAGCCAGTCTGCG 24657
Qy 735 -----AlaIlePheSerLysGluThrGln----- 743
Dh 24658 GCTTCCACGTATGGAGCAACAAGTGAAGTGGAAAGTCTGCCACACTACACAACAAGTGC 24717
Qy 744 -----PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Dh 24718 CATGCTATTTTCAACAGATCATGAGACGCTTGCCTTAACATGACACCA---AAGTGT 24774
Qy 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Dh 24775 GAGCATCAGCGCACAAAGATTTGTGACTGTGGTGCATTTGCTCAATATATGCCACAGGCC 24834
Qy 781 PheAlaIleProTrpLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Dh 24835 TTCTTAAGCCCAATTAAGATGATGATGCGGACGAATTTGCAAAAGCTGCTTATCAAA 24894
Qy 801 AspArgLeuProGlyLysLysThrThrLysLeu-----TrpValProAsp 815
Dh 24895 GAA---GTTCTGGGCAACGAGACTAGCAACTGCAAGACAGACATGATGTTGTGTCGCA 24951
Qy 816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
Dh 24952 GAGAACTACGCGCGGACACTCTATGTAAGTGTATGTCGCTCAAGGCTTATAGCGCAGGTG 25011
Qy 836 LeuLeuGlyMetLysAsnAsnHisSerLysSerLysThrThrIleAlaGlyLeuLeuThr 855
Dh 25012 CTTTGGGATTTGCTTACCGAT---GAGACAGCTCCCAAGAAATTTCCGAATGCTGGCG 25068
Qy 856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
Dh 25069 GCGTTTGTTAACACAGAGGAGATTTGCTGGCCAGAACCGTCTTTCGCGTCCGAGAAA 25128


```

/db_xref="taxon:7227"
/chromosome="2R"
/map="49a-49b"
/clone="BACR14L19 (D591)"
/clone.lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcORI in
PBACE3.6)"
BASE COUNT      54505 a 41459 c 42141 g 54658 t
ORIGIN

Alignment Scores:
Pred. No.:      2,56e-80      length:      192763
Score:          1729.50      Matches:      463
Percent Similarity: 49.43%      Conservative: 270
Best Local Similarity: 31.22%      Mismatches: 436
Query Match:    24.04%      Indels:      318
DB:             3           Gaps:          37

US-09-512-581b-2 (1-1391) x AC007474 (1-192763)

QY      13  ILeThrTYrProProGlyValLysGluIleSerAspLysIleSerLysGluMetVal 32
      146305 ATAGGTACCCACACCGGATGCGCGCGCTGTGAGATTTGGGACAGACGACGACATA 146364
      33  ArgArgLeuLysMet-ValVal-----
      146365 AGGCGCTCAAGGTAGTCTCCGCCCTGTCCCTAGCAAACTCTCACTTAGAATGA 146424
      40  -----LysThrPheMetAspMetAspGlnAspSerGluGlu 52
      146425 TTCTGTTTACAGACCTCCGCAATGTTCTGCAGACCATGACGACGACGACAAATCTCTA 146484
      QY      52  uLysGluLeuTYrLeuLysLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPr 72
      146485 CCAGCA--TACATACCATTTGGCGCCCTCCATCTCGTCGAGACACTTTTATGCAACATCC 146541
      QY      72  oGlyLysAspValArgLeuLeuValAlaCysLysLeuAlaAspIlePheArgIleTYrAl 92
      146542 ATCCCGTGATGCTCCAGCTCTTAAACGCGCTGTGCGGATGCTCCGCGGGTTTATGC 146601
      Db      92  aProGluAlaProTYrThrSerProAspLysLeuLysAspIlePheMetPheIleThrAr 112
      146602 GCCAGAGGCTCCCTACAAAGAGCAAGTCAATCAATCAATATTTAAGTTTATTATCAA 146661
      QY      112  gGlnLeuLysGlyLeuGlnAspThrLysSerProGlnPheAsnArgTYrPheTYrLeuLe 132
      146662 GCAACTGCAGGCTGTGAAGATCGCGCGATCCCTCTTCAAGCGTTACTTCTATGCT 146721
      QY      132  uGluAsnIleAlaTYrValLysSerTYrAsnIleCysPheGluLeuGlnAspSerAsnGl 152
      146722 GGAGAACCTTCTTGTCTGCAAGCTTTTAAATATGTGTTTGAACCTGGAGGACTGCCAAGA 146781
      Db      152  uIlePheThrGlnLeuTYrArgThrPheAspSerValIle----- 165
      146782 AATCTTCCAAGACCTGTTTGTAGTACCATCTTCAGATTGTCAAGTAAGTCCAGCTTCTT 146841
      QY      166  -----AsnAsnGlnHisAsnGl 171
      146842 TTATCTCGCACTTTTATATACATGTTATATATTTTCTAGTACACGACGACATGT 146901
      Db      171  nLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGlnLysAspThrVa 191
      146902 CAAGGTCACTAACTTTTCTAGTGTGTGAAGCCCTGTGATCACAAGACCCGCAATCT 146961
      QY      191  lSerGlnGluLeuLysAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
      146962 GTGCGGTGAATGTTGATCTCATCTGATTAACTGTAGAGCCGTAACAAATCTAACAA 147021
      Db      211  nLysGlnAlaTYrAspLeuAlaLysAlaLeuLeuLysArgTYrHisLagIleAlaIleGlu-P 231
      147022 CAAGTTTGCTGCCAATTAACAGAACAACTTCTCACAATAAACGGGCGATCTCTCGAGTC 147081
      QY      231  ro-----TYrIleThr----- 235

```

```

      147082 CACCATCAAAATGTAATATTTCGTGATCTTTGGTCACACAGTTATTCACCTAATCC 147141
      QY      236  -----PhePheAsnGlnValLeuMetLeuGlyLys-----ThrSerIleSerA 250
      147142 CTGATCTACAGCTTTTAAACCGCTGTGTGTGATGAAGCCCAACACAAACATGCTCC- 147200
      QY      250  spleuSerGlnHisValPheAspLeuIleLeuGluLeuTYrAsnIleAspSerHisLeuL 270
      147201 -ATTACAAACAAATATGATATCATTTAGAGGCTTAACCCGATCAATGACGGCTTGT 147258
      Db      270  euLeuSerValLeuProGlnLeuGlnLupheLysLeuLysSerAsnAspAsnGluArgL 290
      147259 TGTGTGAGTCACTCACTGAGAACAACTGTCTCCAGGAGATGCTGAGAGGC 147318
      QY      290  eu----- 290
      147319 TAACTGAGTAATTTATACAGATGACCTCATTTTGCATATTAATTAACATATTTCTGTTA 147378
      Db      291  -----GlnValValLysLeuLeuAlaLysMetPheGluAlaLysAspSerGluLeuAla 309
      147379 CTACAGAGGCAACACACTCTGTCTGTGATGTTCTCGAAAGAGACTCGCAACTAGCGA 147438
      QY      309  erGlnAsnLysProLeuTYrGlnCysTYrLeuGlyArgPheAsnAspIleHisValProI 329
      147439 AAAAGTACCCCAACCTGTGAAAATCTTTGGCGCTTCTCGACATCACCGAACACAG 147498
      QY      329  leArgLeuGlnCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAla 349
      147499 TTGCATCAACAGTGTGTTCACTCATCCATGACCTTCGTCATACCCCAAGCTTTCAGC 147558
      Db      349  ysAspLeuThrGluTYrLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369
      147559 ACGATTAACACGAGAAATGTGCGGCTCGAATATACAGATCTTGACGAGGTGTCGCCACG 147618
      QY      369  spValIleValSerIleValIleThrAlaLysLysAspIleLeuLeuValAsnAsp---- 387
      147619 AGGGTGTAATAGGCTATGTGTGAAATCCCAAGCGCGCACTCCAGCTGCTCGAAGCCG 147678
      Db      388  -HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTYrArgValArgLysG 407
      147679 CCGATCTACTAGAAATGTGTGCGCGCGGACGCTAGATTAAGATCAAAATTCGCGAGGG 147738
      QY      407  lAlaLeuMetGlyLeuAlaGlnIleTYrLysLysTYrAlaLeuGlnSerAlaAlaGlyL 427
      147739 ATCGATGATAGTGTGCTGCTTACATCAAGCGCGCAATTTGCAACCAAGATCTAA 147798
      QY      427  ysAspAlaAlaLys-----GlnIleAlaTYrPheLysAspLysLeuLeuHisIleTYrTG 446
      147799 GCACTGCGCTTAAGGTGAGGGGTGACTGTAAAGAACAGATACATGATGATACATA 147858
      QY      446  lAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTYrMetValProH 466
      147859 AAGTGGCTTGGAGAGACCGCTGCTGTGAAAGCGCTACTTATACCTGCTGGTCCCT 147918
      Db      466  lAsnLeuGluTYrThrGluArgMetLysCysLeuTYrTYrLeuTYrAlaThrLeuAspL 486
      147919 ATAACTAGCTCCGAGAGAGCGCATGAAGCATATATCAATTTGTGAGGATCTGATG 147978
      QY      486  euAsnAlaValLysAlaLeuAsnGluMetTYrPheLysCysGlnAsnLeuLeuArgHisGln 506
      147979 CCAATGCGACCAAGGCTTTGTGTGAGCTGCGAGAACCAAAATGAACACGCAATPACG 148038
      QY      506  alLysAspLeuLeuAspLeuIleLysGln-----ProLysThrAspAlaSerV 522
      148039 TCAGTGATTTGGATCAAGCTACACCACTTAAGGATTTCAAGCGCGGCTTACTTAGTACG 148098
      Db      522  alLysAla----- 524
      148099 TCAGGCGCAAGCAGCGCAACATTTGCCAAGTGGGTTCTATAATAAGTTTCTTTGATAGT 148158
      QY      525  -IlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543
      525  ---IlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543

```

Db	148159	TTGTACACAGTGAATGTTTTTTTTTTCTTCCACACAGCTCCAGATCCCTTAAGGGGCA	148218
Oy	544	AspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArgLysGlnLeu	563
Db	148219	GAGTACCTAACCCACATTATTAGCAACAACCGCGCAAAAAGATGCATCAGCTCTACAGATGCATT	148278
Oy	564	GluValIleValLysSerProThrCysSerCysLysGlnIleGluGlyCysValArgGluIle	583
Db	148279	AACATTGTCCTTAAACGGCAGCTAAGCTGTGCGGAGTGTGTCAGACACGATGGGAGTTTTC	148338
Oy	584	ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGlnIleMetIleLysPhe	603
Db	148339	CTGAAAAAGCTTTGGCGCCCGCCAGCTCCAA---TCGATCTCTATTACACACAGCTTAAGATG	148395
Oy	604	LeuLeuGlnArgIleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLys	623
Db	148396	CTGATTGAGCGCGTGGCATCATGATGTGGACAAAGAGCTCATTTGGCGCTACTTATTAAAG	148455
Oy	624	-----Gln	624
Db	148456	TAAATTCGTTTAACTGCCCTTTGCCAGACTTAATCCGCTAATCTTTTGGCTACAGCTTA	148515
Oy	625	ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGln	644
Db	148516	ATTGAACCATGATGATCGAGAAAAGGCTCATGTGTGAAGAAGATGGAAATTTGGCGTCAAGAA	148575
Oy	645	AlaIleArgAlaGlyLeuGlnLeuLeu-----	653
Db	148576	GCGGGCGAGCGCGGAGCTCAAGTTGTTGGCCCGCTGAGTATGAGATTAATTAAAGTTATGATA	148635
Oy	654	-----LysValLeuSerPheThrHisProIleSer	663
Db	148636	CAGAAATCCATTGATTCCTTTATCTTCGCAAGATG-CTTTCATACGCTCTTTTCGCGTCAAT	148694
Oy	664	PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLys	683
Db	148695	TTCTTCACCGATCTCTCGCTGCGTATTAATTTCCCTGCTCAGTTACAGCAGGAGTTTAT	148754
Oy	684	ValAlaGlnAlaAlaLeuGlnIlePheLysAsnThrGly-----SerLysIleGln	700
Db	148755	GTTGCACCACTGCTCTAAAGACACTAATCTATCTGGGGCGCTATCAGCCACTCATTAAC	148814
Oy	701	GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLysLys	720
Db	148815	GACCTTACACCGGCTATTCTCGACAGCGTGGCTCGGATGTGCAAGACACTTTCGCACTTAT	148874
Oy	721	GlyProPheArgGlnAlaLysTyrAlaIleHisCysIleHis-----	734
Db	148875	GGAAACCCCAAGCAAGCAAAAGCAGCGGCGATGATTTTTCGTAACAGCCAGTCGTGC	148934
Oy	735	-----AlaIlePheSerLysGluThrGln-----	743
Db	148935	GCTTCCACTGATGGAGCAACAAGTGGACTGGAAATGCTGCACAACTACACAACACATG	148994
Oy	744	-----PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu	760
Db	148995	CATCCTATTTCACAGAGATATCATCAGACGCTTCCGCTTAACACTGCACCA---AAGCTG	149051
Oy	761	GluHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuLeuAlaProAspGln	780
Db	149052	GAGCATTCGCGCACCAAAAGATTTGTGACTTTGGGTGCACATTCCTACAAATATGCGCAGGCC	149111
Oy	781	PheAlaIleAlaProThrPlySerTyrValAlaThrPheIleValLysAspLeuLeuMetAsn	800
Db	149112	TTCTCAACGCCCATTAAGAAATATGATGTGGCGACAGCAATTTGTCAAAAGCTGCTTATCCAA	149171
Oy	801	AspArgLeuProGlyLysLysThrThrLysLeu-----TyrValProAsp	815
Db	149172	GAA---GTTCTCGGCCACGACGACATCAGCAACTGCGCAGAGAGACAGTGAATGTTGGTGGCCAA	149228
Oy	816	GluGluValIleSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyr	835
Db	149229	GAGAAACATACGCGCGACACACTTATAGTAACTGGATGTGGCTTCAAGAGCTTGGCCAGGTGG	149288

QY	836	leuLeuGIyMeLysAsnAsnHisSerLysSerGIyThrSerThrLeuLeuArgLeuThr	855
Db	149289	CTTTGGGATTCGCGAACCGAT---GAGCAGCGCGCCGACAAAACATTTCCGAATGCTGGCG	149345
QY	856	ThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMet	875
Db	149346	GCGTTTGTTAAACCAACGAGCGATTTGCTCGCGCCACAAACCGTCTTTGGCTGGCCGAGAAA	149405
QY	876	SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyr	895
Db	149406	TCTTGGCTGGCGCTGGCGGACCGCTGGCGCATGCTCAAAAGTGTGCCAACAAGAGCGCGTA	149465
QY	896	HisGluIleIleThrLeuGlnGlnTyr---GlnLeuCysAlaIleValIleAsnAspGlu	914
Db	149466	GGTGTACAGTACAGCGCGCTGAGCAGTATTGGCAGCTTTCCAGCTGATGGTGAATACAGAC	149525
QY	914	-----	914
Db	149526	ATTCACTCTTGCTAGAACCTAGTACTCAATGCTCATGCTCGGCGTTCCCGTAGCGTATGC	149588
QY	915	---CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu	933
Db	149586	CGGTGCCA-GAAGTTCGGGAATCTTTGCTCCCAAGCTCCACAAAGATTAACAGAAAGT	149644
QY	934	Arg-----LeuProLeuGlnLysIleMetAlaIleCysAlaLeuCysAla	947
Db	149645	TTGCCCAAGAACTGTTGGCGCTGGACCTTCAATGGCGTTGATGTGGCTGGCTGATAGAG	149704
QY	948	-----LysAspProValLysGlnArgAlaGlnAlaHisAlaArg-GlnCysLeuValLys	964
Db	149705	ACTGAGAGCAAGTGAAGTGAATTCAGAAACCATCCGATCCGTTGGGAAGTATTAATAA	149764
QY	964	S-----AsnIleAsnValArg	969
Db	149765	TATTTTCAACAGATTGCCAAGAACTTGTGGCTGACTATGCGAAGAACGATTAACAAGAC	149824
QY	969	GArgGluTyrLeuLys-	974
Db	149825	GCGGGAATATGCTCAAGAACTGTGGCTATGACATGTAGGTGTTTACTTTTGAATCATATAT	149884
QY	975	-----GlnHis-AlaAlaValSerGlnLysL	983
Db	149885	GTCGATTTTAAACCATTAATCAATATATTCCAACTCCGACAACTCAACGAGTCAACAAT	149944
QY	983	euleSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAsp	1003
Db	149945	CATTACACATACTACTGACTGACTACATAGCTGGCTTTGCTGATTTCCCGTGGCTGGTCCAGATC	150004
QY	1003	roAspTyrValLysValGlnAspIleGlnGlnLeuLysAspValLysGlnCysLeuTrp	1023
Db	150005	CAGCGTTACAGAAATCAGGAGCACTGCTACACTGTGGCAAGATGGGAAGTGGCTCGT	150064
QY	1023	heValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysM	1043
Db	150065	TCATTCTGAGCGCGCTGATGGCCAAACGAAACGTTTGTCCATACCTTCTACAAACGAC	150124
QY	1043	etValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn-	1062
Db	150125	TGCTGACACTGATTAAGATCCGAGATTCAGTCTGGGG---TCGGCAAGCGCGACAACT	150181
QY	1062	-----	1062
Db	150182	ATGATATGCACTCAATCATTTTCCAGCGCTTTCCTCTCAAAATTATATTATACATATAT	150241
QY	1063	-----GluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerL	1079
Db	150242	TTTTGGTATTCAGAAATATGGGGCGCTGTGGCATCTTCCATAGTACATTATGATGATCCA	150301
QY	1079	ys-----SerThrThrTyrSerIleuGluSerProLysAspProV	1092
Db	150302	AGTTTCAGTCCATTGTGATGGCAACAGAGCAAGCTTTTCAATGGCGCTG-----G	150349

```
QY 1092 allLeuProAlaArgPheThrGlnProAsp---LysAsnPheSerAsnThrLysAsnT 1111
      |||||
      ::::: |||||
Db 150350 CTTGCCAGAAATGATTAATAAGAGCGCTCGGTGGGAATTTCCAAACAAATGACGCT 150409
      |||||
QY 1111 yrlLeuProGluMetLysSerPheThrProGlyLysProLysThrThrAsnValL 1131
      ||::|||
      :::::
Db 150410 ATATACCGCTGGAGCTG-----TATACGCTGGAGGCCAAATCCAGCAGCAAGCTG 150460
      |||||
QY 1131 euGLyAlaValAsnLysProLeu---SerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
      |||||
      ::::: |||||
Db 150461 CCGCAACAGC-AATGACAAAGCTGCGCAGCAGAGTGCTC-CAAAGAGACCGCGCGACA 150518
      |||||
QY 1150 erArg-----MetGluThrValSerAsnAlaSerSerS 1161
      ||::|
      ::::: |||||
Db 150519 GTCAATCATGATGATGAAATCCGACAGTGAAGTGAAGTCTTGAAATATCTGAGCAATG 150578
      |||||
QY 1161 erSerAsnProSerSerProGlyArgLLeuLysGlyArgLeuAspSerSerGluMetAspH 1181
      |||||
      ::::: |||||
Db 150579 AATCTAAGTGTGTGAAATTTGANTTTCAGGAGAACAAATCTGTTCACACATAC- 150634
      |||||
QY 1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
      |||||
      ::::: |||||
Db 150635 -----GAGCGGCTGACACTACGAGAGCCCATGCCCCAAGAGAACGCGCGCAGGAG 150682
```

Search completed: September 25, 2003, 00:10:10
Job time : 12397 secs